DOCUMENT RESUME

ED 444 127	CS 014 073
TITLE	Teaching Children To Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Reports of the Subgroups.
INSTITUTION	National Reading Panel, Bethesda, MD.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
REPORT NO	NIH-00-4754
PUB DATE	2000-04-00
NOTE	506p.; See CS 014 072-075 for other "Teaching Children To Read" documents.
AVAILABLE FROM	Web site: http://www.NationalReadingPanel.org/.
PUB TYPE	Information Analyses (070) -- Reports - Descriptive (141)
EDRS PRICE	MF02/PC21 Plus Postage.
DESCRIPTORS	Educational Technology; Elementary Education; *Instructional Effectiveness; Literature Reviews; *Phonics; *Reading Comprehension; *Reading Instruction; *Reading Research; Research Needs; Teacher Education
IDENTIFIERS	Phonemic Awareness; *Reading Fluency

ABSTRACT

This report presents the reports of the subgroups of the National Reading Panel, charged by Congress to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read. Each of the chapters 2 through 6 contains its own executive summary and appendixes in addition to the report of the subgroup. Chapter 1, Introduction and Methodology, discusses the processes applied to the selection, review, and analysis of research relevant to reading instruction. Chapter 2, Alphabetics, contains sections on phonemic awareness instruction and phonics instruction. Chapter 3 addresses reading fluency. Chapter 4 addresses comprehension (containing sections on vocabulary instruction, text comprehension instruction, and teacher preparation and comprehension strategies instruction). Chapter 5 presents the report of the subgroup on teacher education and reading instruction. Chapter 6 addresses computer technology and reading instruction. The report concludes with a minority view that argues the report of the panel neither responds to its charge nor meets the needs of America's schools. (RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 444 127

REPORT OF THE

National Reading Panel

TEACHING CHILDREN TO READ

An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

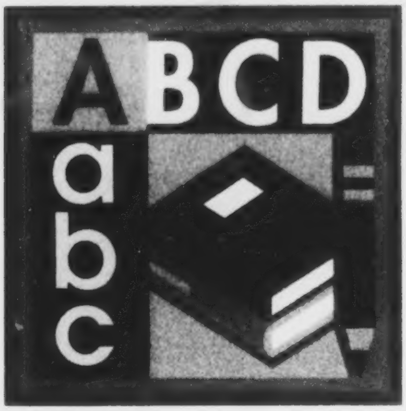
Reports of the Subgroups

2

BEST COPY AVAILABLE

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT
National Institutes of Health

CS 014 073



National Reading Panel

TEACHING CHILDREN TO READ:

An Evidence-Based Assessment
of the Scientific Research Literature
on Reading and Its Implications
for Reading Instruction

REPORTS OF THE SUBGROUPS

Acknowledgments

The National Reading Panel wishes to express its gratitude to the following individuals for their contributions to its effort.

Marilyn Adams
Ed Bouchard
Harris Cooper
Gerald Duffy
Michelle Eidlitz
Barbara Foorman
David Francis
Ester Halberstam
Blair Johnson
Alisa Kenny
Helen S. Kim
Marjolaine Limbos
Khalil Nourani
Simone Nunes
Elizabeth S. Pang
Joan Pagnucco
Michael Pressley
David Reinking
Scott J. Ross
Barbara Schuster
Robin Sidhu
Steven Stahl
Maggie Toplak
Zoreh Yaghoubzadeh

Members of the National Reading Panel

Donald N. Langenberg, Ph.D., *Chair*

Gloria Correro, Ed.D.
Gwenette Ferguson, M.Ed.
Michael L. Kamil, Ph.D.
S.J. Samuels, Ed.D.
Sally E. Shaywitz, M.D.
Joanna Williams, Ph.D.
Joanne Yatvin, Ph.D.

Linnea Ehri, Ph.D.
Norma Garza, C.P.A.
Cora Bagley Marrett, Ph.D.
Timothy Shanahan, Ph.D.
Thomas Trabasso, Ph.D.
Dale Willows, Ph.D.

MEMBERS OF THE NATIONAL READING PANEL SUBGROUPS

Alphabetics
Linnea Ehri, *Chair*
Gloria Correro
Timothy Shanahan
Dale Willows
Joanne Yatvin

Comprehension
Michael L. Kamil, *Chair*
Gwenette Ferguson
Norma Garza
Thomas Trabasso
Joanna Williams

Fluency
S.J. Samuels, *Co-Chair*
Timothy Shanahan, *Co-Chair*
Sally E. Shaywitz

Methodology
Timothy Shanahan, *Co-Chair*
Sally E. Shaywitz, *Co-Chair*

Teacher Education
Gloria Corerro, *Co-Chair*
Michael L. Kamil, *Co-Chair*
Gwenette Ferguson
Norma Garza
Cora Bagley Marrett

Technology/Next Steps
Michael L. Kamil, *Chair*
Donald N. Langenberg

STAFF OF THE NATIONAL READING PANEL

F. William Dommel, Jr., J.D., *Executive Director*

Vinita Chhabra, M.Ed., *Research Scientist*
Judith Rothenberg, *Secretary*
Jaimee Nusbacher, *Meeting Manager*

Mary E. McCarthy, Ph.D., *Senior Staff Psychologist*
Stephanne Player, *Support Staff*
Patrick Riccards, *Senior Advisor*

Table of Contents

Acknowledgments ii

Members of the National Reading Panel iii

Chapter 1: Introduction and Methodology

Introduction 1-1

Methodology: Processes Applied to the Selection, Review, and Analysis of Research Relevant to Reading Instruction 1-5

Chapter 2: Alphabetics

Part I: Phonemic Awareness Instruction

- Executive Summary 2-1
- Report 2-9
- Appendices 2-53

Part II: Phonics Instruction

- Executive Summary 2-89
- Report 2-99
- Appendices 2-145

Chapter 3: Fluency

Executive Summary 3-1
Report 3-5
Appendices 3-35

Chapter 4: Comprehension

Executive Summary 4-1
Introduction 4-11

Part I. Vocabulary Instruction

- Report 4-15
- Appendices 4-33

Part II: Text Comprehension Instruction

- Report 4-39
- Appendices 4-69

Part III. Teacher Preparation and Comprehension Strategies Instruction

- Report 4-119
- Appendices 4-133

Chapter 5: Teacher Education and Reading Instruction

Executive Summary 5-1
Report 5-3
Appendices 5-19

Chapter 6: Computer Technology and Reading Instruction

Executive Summary 6-1
Report 6-3
Appendices 6-13

Minority View

CHAPTER

1

Introduction and Methodology

REPORTS OF THE SUBGROUPS

Introduction

Congressional Charge

In 1997, Congress asked the "Director of the National Institute of Child Health and Human Development (NICHD), in consultation with the Secretary of Education, to convene a national panel to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read." The panel was charged with providing a report that "should present the panel's conclusions, an indication of the readiness for application in the classroom of the results of this research, and, if appropriate, a strategy for rapidly disseminating this information to facilitate effective reading instruction in the schools. If found warranted, the panel should also recommend a plan for additional research regarding early reading development and instruction."

Establishment of the National Reading Panel

In response to this Congressional request, the Director of NICHD, in consultation with the Secretary of Education, constituted and charged a National Reading Panel (the NRP or the Panel). The NRP was composed of 14 individuals, including (as specified by Congress) "leading scientists in reading research, representatives of colleges of education, reading teachers, educational administrators, and parents." The original charge to the NRP asked that a final report be submitted by November 1998.

When the Panel began its work, it quickly became apparent that the Panel could not respond properly to its charge within that time constraint. Permission was sought and received to postpone the report's submission deadline. A progress report was submitted to the Congress in February 1999. The information provided in the NRP Progress Report, the Report of the National Reading Panel, and this Report of the National Reading Panel: Reports of the Subgroups reflects the findings and determinations of the National Reading Panel.

NRP Approach to Achieving the Objectives of Its Charge and Initial Topic Selection

The charge to the NRP took into account the foundational work of the National Research Council (NRC) Committee on Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998). The NRC report is a consensus document based on the best judgments of a diverse group of experts in reading research and reading instruction. The NRC Committee identified and summarized research literature relevant to the critical skills, environments, and early developmental interactions that are instrumental in the acquisition of beginning reading skills. The NRC Committee did not specifically address "how" critical reading skills are most effectively taught and what instructional methods, materials, and approaches are most beneficial for students of varying abilities.

In order to build upon and expand the work of the NRC Committee, the NRP first developed an objective research review methodology. The Panel then applied this methodology to undertake comprehensive, formal, evidence-based analyses of the experimental and quasi-experimental research literature relevant to a set of selected topics judged to be of central importance in teaching children to read. An examination of a variety of public databases by Panel staff revealed that approximately 100,000 research studies on reading have been published since 1966, with perhaps another 15,000 appearing before that time. Obviously, it was not possible for a panel of volunteers to examine critically this entire body of research literature. Selection of prioritized topics was necessitated by the large amount of published reading research literature relevant to the Panel's charge to determine the effectiveness of reading instructional methods and approaches. A screening process was, therefore, essential.

The Panel's initial screening task involved selection of the set of topics to be addressed. Recognizing that this selection would require the use of informed judgment, the Panel chose to begin its work by broadening its understanding of reading issues through a thorough analysis of the findings of the NRC report, Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998). Early in its deliberations the Panel made a tentative decision to establish subgroups of its members and to assign to each subgroup one of the major topic areas designated by the NRC Committee as central to learning to read—Alphabetics, Fluency, and Comprehension.

Regional Public Hearings

As part of its information gathering, the Panel publicly announced, planned, and held regional hearings in Chicago, IL (May 29,1998), Portland, OR (June 5, 1998), Houston, TX (June 8, 1998), New York, NY (June 23, 1998), and Jackson, MS (July 9, 1998). The Panel believed that it would not have been possible to accomplish the mandate of Congress without first hearing directly from consumers of this information—teachers, parents, students, and policymakers—about their needs and their understanding of the research. Although the regional hearings were not intended as a substitute for scientific research, the hearings gave the Panel an opportunity to listen to the voices of those who will need to consider implementation of the Panel's findings and determinations. The regional hearings gave members a clearer understanding of the issues important to the public.

As a result of these hearings, the Panel received oral and written testimony from approximately 125 individuals or organizations representing citizens—teachers, parents, students, university faculty, educational policy experts, and scientists—who would be the ultimate users and beneficiaries of the research-derived findings and determinations of the Panel.

At the regional hearings, several key themes were expressed repeatedly:

- The importance of the role of parents and other concerned individuals, especially in providing children with early language and literacy experiences that foster reading development;
- The importance of early identification and intervention for all children at risk for reading failure;
- The importance of phonemic awareness, phonics, and good literature in reading instruction, and the need to develop a clear understanding of how best to integrate different reading approaches to enhance the effectiveness of instruction for all students;
- The need for clear, objective, and scientifically based information on the effectiveness of different types of reading instruction and the need to have such research inform policy and practice;
- The importance of applying the highest standards of scientific evidence to the research review process so that conclusions and determinations are based on findings obtained from experimental studies characterized by methodological rigor with demonstrated reliability, validity, replicability, and applicability;
- The importance of the role of teachers, their professional development, and their interactions and collaborations with researchers, which should be recognized and encouraged; and
- The importance of widely disseminating the information that is developed by the Panel.

Adoption of Topics To Be Studied

Following the regional hearings, the Panel considered, discussed, and debated several dozen possible topic areas and then settled on the following topics for intensive study:

- Alphabetics
  - Phonemic Awareness Instruction
  - Phonics Instruction
- Fluency
- Comprehension
  - Vocabulary Instruction
  - Text Comprehension Instruction
  - Teacher Preparation and Comprehension Strategies Instruction
- Teacher Education and Reading Instruction
- Computer Technology and Reading Instruction

10

In addition, because of the concern voiced by the public at the regional hearings that the highest standards of scientific evidence be applied in the research review process, the methodology subgroup was tasked to develop a research review process including specific review criteria.

Each topic and subtopic became the subject of the work of a subgroup composed of one or more Panel members. Some Panel members served on more than one subgroup. (The full report of each subgroup is included in this volume.) The subgroups formulated seven broad questions to guide their efforts in meeting the Congressional charge of identifying effective instructional reading approaches and determining their readiness for application in the classroom:

1. Does instruction in phonemic awareness improve reading? If so, how is this instruction best provided?
2. Does phonics instruction improve reading achievement? If so, how is this instruction best provided?
3. Does guided repeated oral reading instruction improve fluency and reading comprehension? If so, how is this instruction best provided?
4. Does vocabulary instruction improve reading achievement? If so, how is this instruction best provided?
5. Does comprehension strategy instruction improve reading? If so, how is this instruction best provided?
6. Do programs that increase the amount of children's independent reading improve reading achievement and motivation? If so, how is this instruction best provided?
7. Does teacher education influence how effective teachers are at teaching children to read? If so, how is this instruction best provided?

Each subgroup also generated several subordinate questions to address within each of the major questions. It should be made clear that the Panel did not consider these questions and the instructional issues that they represent to be the *only* topics of importance in learning to read. The Panel's silence on other topics should not be interpreted as indicating that other topics have no importance or that improvement in those areas would not lead to greater reading achievement. It was simply the sheer number of studies identified by Panel staff relevant to reading (more than 100,000 published since 1966 and more than 15,000 prior to 1966) that precluded an exhaustive analysis of the research in all areas of potential interest.

The Panel also did not address issues relevant to second language learning, as this topic was being addressed in detail in a new, comprehensive NICHD/OERI (Office of Educational Research and Improvement) research initiative. The questions presented above bear on instructional topics of widespread interest in the field of reading education that have been articulated in a wide range of theories, research studies, instructional programs, curricula, assessments, and educational policies. The Panel elected to examine these and subordinate questions because they currently reflect the central issues in reading instruction and reading achievement.

REPORTS OF THE SUBGROUPS

Methodology: Processes Applied to the Selection, Review, and Analysis of Research Relevant to Reading Instruction

In an important action critical to its Congressional charge, the NRP elected to develop and adopt a set of rigorous research methodological standards. These standards, which are defined in this section, guided the screening of the research literature relevant to each topic area addressed by the Panel. This screening process identified a final set of experimental or quasi-experimental research studies that were then subjected to detailed analysis. The evidence-based methodological standards adopted by the Panel are essentially those normally used in research studies of the efficacy of interventions in psychological and medical research. These include behaviorally based interventions, medications or medical procedures proposed for use in the fostering of robust health and psychological development and the prevention or treatment of disease.

It is the view of the Panel that the efficacy of materials and methodologies used in the teaching of reading and in the prevention or treatment of reading disabilities should be tested no less rigorously. However, such standards have not been universally accepted or used in reading education research. Unfortunately, only a small fraction of the total reading research literature met the Panel's standards for use in the topic analyses.

With this as background, the Panel understood that criteria had to be developed as it considered which research studies would be eligible for assessment. There were two reasons for determining such guidelines or rules *a priori*. First, the use of common search, selection, analysis, and reporting procedures would ensure that the Panel's efforts could proceed, not as a diverse collection of independent—and possibly uneven—synthesis papers, but as parts of a greater whole. The use of common procedures permitted a more unified presentation of the combined methods and findings. Second, the amount of research synthesis that had to be accomplished was substantial. Consequently, the Panel had to work in diverse subgroups to identify, screen, and evaluate the relevant research to complete their respective reports. Moreover, the Panel also had to arrive at findings that all or nearly all of the members of the NRP could endorse. Common procedures, grounded in scientific principles, helped the Panel to reach final agreements.

Search Procedures

Each subgroup conducted a search of the literature using common procedures, describing in detail the basis and rationale for its topical term selections, the strategies employed for combining terms or delimiting searches, and the search procedures used for each topical area.

Each subgroup limited the period of time covered by its searches on the basis of relative recentness and how much literature the search generated. For example, in some cases it was decided to limit the years searched to the number of most recent years that would identify between 300 to 400 potential sources. This scope could be expanded in later iterations if it appeared that the nature of the research had changed qualitatively over time, if the proportion of useable research identified was small (e.g., less than 25%), or if the search simply represented too limited a proportion of the total set of identifiable studies. Although the number of years searched varied among subgroup topics, decisions regarding the number of years to be searched were made in accord with shared criteria.

The initial criteria were established to focus the efforts of the Panel. First, any study selected had to focus directly on children's reading development from preschool through grade 12. Second, the study had to be published in English in a refereed journal. At a minimum, each subgroup searched both PsycINFO and ERIC databases for studies meeting these initial criteria. Subgroups could, and did, use additional databases when appropriate. Although the use of a minimum of two databases identified duplicate literature, it also afforded the opportunity to expand perspective and locate articles that would not be identifiable through a single database.

Identification of each study selected was documented for the record and each was assigned to one or more members of the subgroup who examined the title and abstract. Based on this examination, the subgroup member(s) determined, if possible at this stage, whether the study addressed issues within the purview of the research questions being investigated. If it did not, the study was excluded and the reason(s) for the exclusion were detailed and documented for the record. If, however, it did address reading instructional issues relevant to the Panel's selected topic areas, the study underwent further examination.

Following initial examination, if the study had not been excluded in accord with the preceding criteria, the full study report was located and examined in detail to determine whether the following criteria were met:

- Study participants must be carefully described (age; demographics; cognitive, academic, and behavioral characteristics).
- Study interventions must be described in sufficient detail to allow for replicability, including how long the interventions lasted and how long the effects lasted.
- Study methods must allow judgments about how instruction fidelity was ensured.
- Studies must include a full description of outcome measures.

These criteria for evaluating research literature are widely accepted by scientists in disciplines involved in medical, behavioral, and social research. The application of these criteria increased the probability that objective, rigorous standards were used and that therefore the information obtained from the studies would contribute to the validity of any conclusions drawn.

If a study did not meet these criteria or could not be located, it was excluded from subgroup analysis and the reason(s) for its exclusion was detailed and documented for the record. If the study was located and met the criteria, the study became one of the subgroup's core working set of studies. The core working sets of studies gathered by the subgroups were then coded as described below and then analyzed to address the questions posed in the introduction and in the charge to the Panel.

If a core set of studies identified by the subgroup was insufficient to answer critical instructional questions, less recent studies were screened for eligibility for, and inclusion in, the core working sets of studies. This second search used the reference lists of all core studies and known literature reviews. This process identified cited studies that could meet the Panel's methodological criteria for inclusion in the subgroups' core working sets of studies. Any second search was described in detail and applied precisely the same search, selection, exclusion, and inclusion criteria and documentation requirements as were applied in the subgroups' initial searches.

Manual searches, again applying precisely the same search, selection, exclusion and inclusion criteria, and documentation requirements as were applied in the subgroups' electronic searches, were also conducted to supplement the electronic database searches. Manual searching of recent journals that publish research on specific NRP subgroup topics was performed to compensate for the delay in appearance of these journal articles in the electronic databases. Other manual searching was carried out in relevant journals to include eligible articles that should have been selected, but were missed in electronic searches.

Source of Publications: The Issue of Refereed and Non-Refereed Articles

The subgroup searches focused exclusively on research that had been published or had been scheduled for publication in refereed (peer reviewed) journals. The Panel reached consensus that determinations and findings for claims and assumptions guiding instructional practice depended on such studies. Any search or review of studies that had not been published through the peer review process but was consulted in any subgroups review was treated as separate and distinct from evidence drawn from peer-reviewed sources (i.e., in an appendix) and is not referenced in the Panel's report. These non-peer-reviewed data were treated as preliminary/pilot data that might illuminate potential trends and areas for future research. Information derived in whole or in part from such studies was not to be represented at the same level of certainty as findings derived from the analysis of refereed articles.

Types of Research Evidence and Breadth of Research Methods Considered

Different types of research (e.g., descriptive-interpretive, correlational, experimental) lay claim to particular warrants, and these warrants differ markedly. The Panel felt that it was important to use a wide range of research but that that research be used in accordance with the purposes and limitations of the various research types.

To make a determination that any instructional practice could be or should be adopted widely to improve reading achievement requires that the belief, assumption, or claim supporting the practice be causally linked to a particular outcome. The highest standard of evidence for such a claim is the experimental study, in which it is shown that treatment can make such changes and effect such outcomes. Sometimes when it is not feasible to do a randomized experiment, a quasi-experimental study is conducted. This type of study provides a standard of evidence that, while not as high, is acceptable, depending on the study design.

To sustain a claim of effectiveness, the Panel felt it necessary that there be experimental or quasi-experimental studies of sufficient size or number, and scope (in terms of population served), and that these studies be of moderate to high quality. When there were either too few studies of this type, or they were too narrowly cast, or they were of marginally acceptable quality, then it was essential that the Panel have substantial correlational or descriptive studies that concurred with the findings if a claim was to be sustained. No claim could be determined on the basis of descriptive or correlational research alone. The use of these procedures increased the possibility of reporting findings with a high degree of internal validity.

Coding of Data

Characteristics and outcomes of each study that met the screening criteria described above were coded and analyzed, unless otherwise authorized by the Panel. The data gathered in these coding forms were the information submitted to the final analyses. The coding was carried out in a systematic and reliable manner.

The various subgroups relied on a common coding form developed by a working group of the Panel's scientist members and modified and endorsed by the Panel. However, some changes could be made to the common form by the various subgroups for addressing different research issues. As coding forms were developed, any changes to the common coding form were shared with and approved by the Panel to ensure consistency across various subgroups.

Unless specifically identified and substantiated as unnecessary or inappropriate by a subgroup and agreed to by the Panel, each form for analyzing studies was coded for the following categories:

1. Reference

- Citation (standard APA format)
- How this paper was found (e.g., search of named database, listed as reference in another empirical paper or review paper, hand search of recent issues of journals)
- Narrative summary that includes distinguishing features of this study

2. Research Question: The general umbrella question that this study addresses.

3. Sample of Student Participants

- States or countries represented in sample
- Number of different schools represented in sample
- Number of different classrooms represented in sample
- Number of participants (total, per group)
- Age
- Grade
- Reading levels of participants (prereading, beginning, intermediate, advanced)

- Whether participants were drawn from urban, suburban, or rural setting
- List any pretests that were administered prior to treatment
- List any special characteristics of participants including the following if relevant:
- Socioeconomic status (SES)
- Ethnicity
- Exceptional learning characteristics, such as:
    - Learning disabled
    - Reading disabled
    - Hearing impaired
- English Language Learners (ELL)—also known as Limited English Proficient (LEP) students
- Explain any selection restrictions that were applied to limit the sample of participants (e.g., only those low in phonemic awareness were included)
- Contextual information: Concurrent reading instruction that participants received in their classrooms during the study
    - Was the classroom curriculum described in the study (code = yes/no)
    - Describe the curriculum
- Describe how sample was obtained:
    - Schools or classrooms or students were selected from the population of those available
    - Convenience or purposive sample
    - Not reported
    - Sample was obtained from another study (specify study)
- Attrition:
    - Number of participants lost per group during the study
    - Was attrition greater for some groups than for others? (yes/no)

4. Setting of the Study

- Classroom
- Laboratory
- Clinic
- Pullout program (e.g., Reading Recovery®)
- Tutorial

5. Design of Study

- Random assignment of participants to treatments (randomized experiment)
    - With vs. without a pretest
- Nonequivalent control group design (quasi-experiment) (Example: existing groups assigned to treatment or control conditions, no random assignment)
    - With vs. without matching or statistical control to address nonequivalence issue
- One-group repeated measure design (i.e., one group receives multiple treatments, considered a quasi-experiment)
    - Treatment components administered in a fixed order vs. order counterbalanced across subgroups of participants
- Multiple baseline (quasi-experiment)
    - Single-subject design
    - Aggregated-subjects design

6. Independent Variables

a. Treatment Variables

- Describe all treatments and control conditions; be sure to describe nature and components of reading instruction provided to control group
- For each treatment, indicate whether instruction was explicitly or implicitly delivered and, if explicit instruction, specify the unit of analysis (sound-symbol; onset/rime; whole word) or specific responses taught. [*Note:* If this category is omitted in the coding of data, justification must be provided.]
- If text is involved in treatments, indicate difficulty level and nature of texts used
- Duration of treatments (given to students)
    - Minutes per session
    - Sessions per week
    - Number of weeks

- Was trainers' fidelity in delivering treatment checked? (yes/no)
- Properties of teachers/trainers
- Number of trainers who administered treatments
- Teacher/student ratio: Number of participants to number of trainers
- Type of trainer (classroom teacher, student teacher, researcher, clinician, special education teacher, parent, peer, other)
- List any special qualifications of trainers
- Length of training given to trainers
- Source of training
- Assignment of trainers to groups:
    - Random
    - Choice/preference of trainer
    - All trainers taught all conditions
- Cost factors: List any features of the training such as special materials or staff development or outside consultants that represent potential costs

b. Moderator Variables

List and describe other nontreatment independent variables included in the analyses of effects (e.g., attributes of participants, properties or types of text).

7. Dependent (Outcome) Variables

- List processes that were taught during training and measured during and at the end of training
- List names of reading outcomes measured
    - Code each as standardized or investigator-constructed measure
    - Code each as quantitative or qualitative measure
    - For each, is there any reason to suspect low reliability? (yes/no)
- List time points when dependent measures were assessed

8. Nonequivalence of Groups

- Any reason to believe that treatment/control group might not have been equivalent prior to treatments? (yes/no)
- Were steps taken in statistical analyses to adjust for any lack of equivalence? (yes/no)

9. Result (for each measure)

- Record the name of the measure
- Record whether the difference—treatment mean minus control mean—is positive or negative
- Record the value of the effect size including its sign (+ or -)
- Record the type of summary statistics from which the effect size was derived
- Record number of people providing the effect size information

10. Coding Information

- Record length of time to code study
- Record name of coder

If text was a variable, the coding indicated what is known about the difficulty level and nature of the texts being used. Any use of special personnel to deliver an intervention, use of special materials, staff development, or other features of the intervention that represent potential cost were noted. Finally, various threats to reliability and internal or external validity (group assignment, teacher assignment, fidelity of treatment, and confounding variables including equivalency of subjects prior to treatment and differential attrition) were coded. Each subgroup also coded additional items deemed appropriate or valuable to the specific question being studied by the subgroup members.

A study could be excluded at the coding stage only if it was found to have so serious a fundamental flaw that its use would be misleading. The reason(s) for exclusion of any such study was detailed and documented for the record. When quasi-experimental studies were selected, it was essential that each study included both pre-treatment and post-treatment evaluations of performance and that there was a comparison group or condition.

Each subgroup conducted an independent re-analysis of a randomly designated 10% sample of studies. Absolute rating agreement was calculated for each category (not for forms). If absolute agreement fell below 0.90 for any category for occurrence or non-occurrence agreement, the subgroup took some action to improve agreement (e.g., multiple readings with resolution, improvements in coding sheet).

Upon completion of the coding for recently published studies, a letter was sent to the first author of the study requesting any missing information. Any information that was provided by authors was added to the database.

After its search, screening, and coding, a subgroup determined whether for a particular question or issue a meaningful meta-analysis could be completed, or whether it was more appropriate to conduct a literature analysis of that issue or question without meta-analysis, incorporating all of the information gained. The full Panel reviewed and approved or modified each decision.

Data Analysis

When appropriate and feasible, effect sizes were calculated for each intervention or condition in experimental and quasi-experimental studies. The subgroups used the standardized mean difference formula as the measure of treatment effect. The formula was:

$(M_t - M_c) / 0.5(sd_t + sd_c)$

where:

M_t is the mean of the treated group,
M_c is the mean of the control group,
sd_t is the standard deviation of the treated group,

and

sd_c is the standard deviation of the control group.

When means and standard deviations were not available, the subgroups followed the guidelines for the calculation of effect sizes as specified in Cooper and Hedges (1994).

The subgroups weighted effect sizes by numbers of subjects in the study or comparison to prevent small studies from overwhelming the effects evident in large studies. Each subgroup used median and/or average effect sizes when a study had multiple comparisons, and only employed the comparisons that were specifically relevant to the questions under review by the subgroup.

Expected Outcomes

Analyses of effect sizes were undertaken with several goals in mind. First, overall effect sizes of related studies were calculated across subgroups to determine the best estimate of a treatment's impact on reading. These overall effects were examined with regard to their difference from zero (i.e., does the treatment have an effect on reading?), strength (i.e., if the treatment has an effect, how large is that effect?), and consistency (i.e., did the effect of the treatment vary significantly from study to study?). Second, the Panel compared the magnitude of a treatment's effect under different methodological conditions, program contexts, program features, outcome measures, and for students with different characteristics. The appropriate moderators of a treatment's impact were drawn from the distinctions in studies recorded on the coding sheets. In each case, a statistical comparison was made to examine the impact of each moderator variable on average effect sizes for each relevant outcome variable. These analyses enabled the Panel to determine the conditions that alter a program's effects and the types of individuals for whom the program is most and least effective. Within-group average effect sizes were examined, as were overall effect sizes, for differences from zero and for strength. The analytic procedures were carried out using the techniques described in Cooper and Hedges (1994).

References

Cooper, H., & Hedges, L.V. (1994). The handbook of research synthesis. New York: Russell Sage Foundation.

Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

CHAPTER

2

Alphabetics

CHAPTER 2

Alphabetics

Part I

Phonemic Awareness Instruction

PART I: PHONEMIC AWARENESS INSTRUCTION

Executive Summary

Introduction

When today's educators discuss the ingredients of effective programs to teach children to read, phonemic awareness (PA) receives much attention. However, not everyone is convinced. In education, particularly in the teaching of reading over the years, the choice of instructional methods has been heavily influenced by many factors, not only teachers' own frontline experiences about what works, but also politics, economics, and the popular wisdom of the day. The pendulum has swung back and forth between holistic, meaning-centered approaches and phonics approaches without much hope of resolving disagreements. Meanwhile, substantial scientific evidence has accumulated purporting to shed light on reading acquisition processes and effective instructional approaches (Anderson et al., 1985; Adams, 1990; Snow, 1998). Many studies investigating the effectiveness of phonemic awareness instruction have contributed to this body of evidence. Proponents believe that this research holds promise of placing reading instruction on a more solid footing and ending the periodic upheavals and overhauls of reading instructional practices.

The purpose of this report of the National Reading Panel (NRP) was to examine the scientific evidence relevant to the impact of phonemic awareness instruction on reading and spelling development. In the analyses conducted, the NRP sought answers to questions such as the following: Is phonemic awareness instruction effective in helping children learn to read? Under what circumstances and for which children is it most effective? Were studies showing its effectiveness designed appropriately to yield scientifically valid findings? What does a careful analysis of the findings reveal? How applicable are these findings to classroom practice? To evaluate the adequacy and strength of the evidence, the NRP conducted a meta-analysis. The literature was searched to locate all experimental studies that included a PA treatment and a control group and that measured reading as an outcome of the treatment.

There were several reasons why phonemic awareness instruction was selected for review and analysis. Correlational studies have identified phonemic awareness and letter knowledge as the two best school-entry predictors of how well children will learn to read during their first 2 years in school. This evidence suggests the potential instructional importance of teaching PA to children. Many experimental studies have evaluated the effectiveness of PA instruction in facilitating reading acquisition. Results are claimed to be positive and to provide a scientific basis documenting the efficacy of PA instruction. There is currently much interest in PA programs among teachers, principals, and publishers. State adoption committees have prescribed the inclusion of PA training in reading instruction materials approved for use in schools. It is thus important to determine whether PA instruction lives up to these claims and, if so, to identify circumstances that govern its effectiveness.

Phonemes are the smallest units constituting spoken language. English consists of about 41 phonemes. Phonemes combine to form syllables and words. A few words have only one phoneme, such as a or oh. Most words consist of a blend of phonemes, such as go with two phonemes, or check with three phonemes, or stop with four phonemes. Phonemes are different from graphemes, which are units of written language and which represent phonemes in the spellings of words. Graphemes may consist of one letter, for example, P, T, K, A, N, or multiple letters, CH, SH, TH, -CK, EA, -IGH, each symbolizing one phoneme.

Phonemic awareness refers to the ability to focus on and manipulate phonemes in spoken words. The following tasks are commonly used to assess children's PA or to improve their PA through instruction and practice:

1. Phoneme isolation, which requires recognizing individual sounds in words, for example, "Tell me the first sound in paste." (/p/)
2. Phoneme identity, which requires recognizing the common sound in different words. For example, "Tell me the sound that is the same in bike, boy, and bell." (/b/)
3. Phoneme categorization, which requires recognizing the word with the odd sound in a sequence of three or four words, for example, "Which word does not belong? bus, bun, rug." (rug)
4. Phoneme blending, which requires listening to a sequence of separately spoken sounds and combining them to form a recognizable word. For example, "What word is /s/ /k/ /u/ /l/?" (school)
5. Phoneme segmentation, which requires breaking a word into its sounds by tapping out or counting the sounds or by pronouncing and positioning a marker for each sound. For example, "How many phonemes are there in ship? " (three: /š/ /I/ /p/)
6. Phoneme deletion, which requires recognizing what word remains when a specified phoneme is removed. For example, "What is smile without the /s/?" (mile)

In the studies reviewed by the NRP, researchers used one or several of these tasks to assess how much PA children possessed before training and how much they had learned at the end of training. Also, these tasks were the basis for activities that children practiced during training. In some of the studies, children were taught to perform these tasks with letters, for example, segmenting words into phonemes and representing each with a grapheme. In other studies, phoneme manipulation was limited to speech.

To be clear, PA instruction is not synonymous with phonics instruction that entails teaching students how to use grapheme-phoneme correspondences to decode or spell words. PA instruction does not qualify as phonics instruction when it teaches children to manipulate phonemes in speech, but it does qualify when it teaches children to segment or blend phonemes with letters.

PA is thought to contribute to helping children learn to read because the structure of the English writing system is alphabetic. Moreover, it is not easy to figure out the system. Although most English words have prescribed spellings that consist of graphemes, symbolizing phonemes in predictable ways, being able to distinguish the separate phonemes in pronunciations of words so that they can be matched to graphemes is difficult. This is because spoken language is seamless; that is, there are no breaks in speech signaling where one phoneme ends and the next one begins. Rather, phonemes are folded into each other and are coarticulated. Discovering phonemic units requires instruction to learn how the system works.

Methodology

How was the analysis of the research literature conducted?

Before conducting a meta-analysis, the NRP systematically searched the research literature relevant to PA instruction. After a methodology established by the Panel was followed, appropriate key words were entered to identify relevant studies in ERIC and PsycINFO. The search was limited to articles appearing in journals written in English, but no limit was placed on the year of publication. This yielded a total of 1,962 potentially relevant articles. Abstracts were printed and screened. In addition, references listed in these articles and in several review papers were hand-searched and screened. To qualify for analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group or a multiple baseline method.
2. Studies had to appear in a refereed journal.
3. Studies had to test the hypothesis that instruction in phonemic awareness improves reading performance over alternative forms of instruction or no instruction.
4. Studies had to provide training in phonemic awareness that was not confounded with other instructional methods or activities.
5. Studies had to report statistics permitting the calculation or estimation of effect sizes.

Applying these procedures, the NRP found 52 articles from which 96 instructional comparisons were drawn. In each comparison, one group of children was taught PA while a control group received either another type of instruction or regular classroom instruction. Following training, the two groups were compared in their ability to read.

The primary statistic used in the NRP analysis was "effect size," the extent to which performance of the treatment group exceeded performance of the control group. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of PA instruction. An effect size of 0 indicates that treatment and control group means were identical, revealing that training had no effect. To judge the strength of an effect size, a value of 0.20 is considered small, 0.50 is moderate, and 0.80 is large. For each comparison, three effect sizes were calculated to determine whether PA instruction improved children's phonemic awareness, reading, and spelling.

The studies in the NRP database varied in many respects. These variations showed whether effect sizes were bigger under some conditions than others. The NRP compared effect sizes associated with the following variations:

- Type of test: a standardized test was used or a test devised by experimenters.
- Time of test: Outcomes were measured right after instruction or after a delay.
- Type of PA training: Children received instruction that focused on one type of PA or two types of PA, or they were taught three or more types of PA skills.
- Use of letters: Children were taught to manipulate phonemes using letters, or they were taught to manipulate phonemes in speech only.
- Size of groups: Children were taught individually or in small groups or in larger classroom groups.
- Trainer: The source of the instruction was the children's classroom teacher or a researcher or a computer.
- Length of instruction: Instruction varied from 1 hour to 75 hours.
- Reading level of students: The children receiving instruction were at risk for developing reading problems, or were reading disabled, or were normally developing readers.
- Grade level: The children were preschoolers, kindergartners, 1st graders, or 2nd through 6th graders.
- Socioeconomic status (SES): The children were low SES or middle-to-high SES.

In addition, the NRP examined various features of the experiments to determine whether those showing strong effects were well designed or weakly designed. Among the design features examined were whether children were randomly assigned to treatment and control groups, whether the size of the sample was small or large, and whether the study met criteria of rigor specified in a critique by Troia (1999).

Results and Discussion

What do results of the meta-analysis of PA instruction studies show?

The NRP examined whether PA instruction was significantly better than alternative forms of training in helping children acquire phonemic awareness and enabling them to apply this skill in their reading and spelling. Results were positive. The overall effect size on PA outcomes was large, 0.86. The overall effect size on reading outcomes was moderate, 0.53. The overall effect on spelling was also moderate, 0.59. Effects were significant on followup tests given several months after training ended. Effects were significant on measures of children's ability to read words and pseudowords as well as their reading comprehension. Effects were significant on standardized tests as well as experimenter-devised tests. These findings show that teaching children to manipulate phonemes in words was highly effective across all the literacy domains and outcomes. Effects of training did not generalize to performance on math tests, indicating that halo/Hawthorne effects did not account for the findings.

What were the effects of moderators on learning phonemic awareness?

The NRP examined whether PA training was effective under more specific conditions. Children acquired PA successfully under all conditions, but some conditions produced larger effects than others. Effect sizes were larger when children received focused and explicit instruction on one or two PA skills than when they were taught a combination of three or more PA skills. Instruction that taught phoneme manipulation with letters helped normally developing readers and at-risk readers acquire PA better than PA instruction without letters. When children were taught PA in small groups, their learning was greater than when they were taught individually or in classrooms. The length of time spent teaching children was influential, with treatments lasting from 5 to 18 hours producing larger effect sizes than shorter or longer treatments. Classroom teachers were very effective in teaching PA to children. Also, computers were effective. Although all levels of readers acquired PA successfully, effect sizes were greater for children who were beginning readers at risk for reading failure and normally progressing readers than for older disabled readers. Students in the lower grades, preschool, and kindergarten, showed larger effect sizes in acquiring PA than children in 1st grade and above. Children learning to read in English showed larger effects than children learning to read in other alphabetic languages. However, SES level exerted no impact on effect size, indicating that low and mid-to-high SES children benefited similarly from PA training in acquiring phonemic awareness.

What were the effects of moderators on learning to read?

The impact of these specific conditions on the amount of transfer from PA training to other reading skills was also examined. For example, transfer was greater when experimenter-devised tests were used to measure reading skills than when standardized tests were used. This was not surprising, given that standardized tests tend to be less sensitive. Teaching that focused on one or two types of PA manipulations yielded larger effect sizes than teaching three or more PA skills. Teaching children to manipulate phonemes using letters produced bigger effects than teaching without letters. Blending and segmenting instruction exerted a significantly larger effect on reading development than did multiple-skill instruction. Small-group instruction produced larger effect sizes on reading than individual instruction or classroom instruction, albeit in an unanticipated fashion. Specifically, the longer the training program, the smaller the effect size. Significant improvement in reading skills following PA instruction was observed both in studies involving classroom teachers and in computer formats, but the degree of transfer was less than that achieved in experimentally controlled studies. Large effect sizes were obtained in studies of at-risk readers, with moderate effect sizes obtained for disabled and normally developing readers.

Moreover, preschoolers exhibited a much larger effect size on reading than did students in the other grade levels. Children learning to read in English also showed larger transfer effects to reading than children learning in other languages. The effects of PA training on reading outcomes were also influenced by SES, with mid-to-high SES associated with larger effect sizes than low SES.

What were the effects of moderators on learning to spell?

The NRP also examined how different conditions influenced the impact and transfer of PA training to spelling. The effects of PA training on spelling for disabled readers was minimal, as indicated by effect sizes that did not differ significantly from zero. This is consistent with other findings indicating that learning to spell is especially difficult for disabled readers. Because disabled readers were unevenly distributed across the conditions that were examined in relation to the effects of PA training on spelling, along with the finding of a nonsignificant effect size, data obtained from studies of disabled readers were eliminated from the database.

The effects of conditions on spelling outcomes were analyzed for at-risk and normal readers. For these groups, effect sizes involving spelling outcomes did not differ across levels of the following properties of PA training: whether one or two or multiple PA skills were taught, whether training was conducted with individuals or small groups or classroom-size groups, how long training lasted, or whether the trainer was a classroom teacher or a researcher. However, effect sizes did differ across other conditions. Teaching children to manipulate phonemes with letters exerted a much larger impact on spelling than teaching children without letters.

Also kindergartners made greater gains from PA training in spelling than 1st graders. Mid-to-high SES children showed larger effect sizes on spelling than low SES children. Children acquiring literacy in English showed larger effects on spelling than children acquiring literacy in other languages.

Did the effects of PA training arise from well-designed experiments?

The NRP examined whether significant effect sizes arose primarily from experiments with the weakest designs or whether well-designed experiments showed significant effect sizes as well. Findings indicated that rigorous designs yielded strong effects. The majority of studies used random assignment, and their effect sizes on PA and reading outcomes ranged from moderate to large. About one-third of the studies assessed trainers' fidelity to instructional procedures. Effect sizes in these studies were moderate.

Some studies compared PA treatment groups to control groups that were given another treatment, and some studies used untreated control groups. Neither type of control group consistently produced larger effect sizes, indicating that Hawthorne effects do not explain why PA training was effective. Although studies using smaller samples tended to show somewhat larger effect sizes, even those having the largest samples showed positive and significant effects that were moderate in size.

The NRP also assessed the relationship between methodological rigor and effect size by applying Troia's (1999) criteria to the studies. On PA outcomes, studies that met his criteria for the best designs produced the largest effect sizes on all five measures of rigor. On reading outcomes, effect sizes associated with the most rigorous levels were close to the largest, if not the largest, effect sizes on four out of five measures. Thus, these findings indicate that claims about the effectiveness of PA instruction are supported by evidence derived from methodologically sound studies.

Conclusions

What conclusions can be drawn from this meta-analysis of PA instruction studies?

Can phonemic awareness be taught?

Yes. The results clearly showed that PA instruction is effective in teaching children to attend to and manipulate speech sounds in words. Findings of the meta-analysis revealed not only that PA can be taught but also that PA instruction is effective under a variety of teaching conditions with a variety of learners.

Does phonemic awareness instruction assist children in learning to read? If so, which students benefit?

Yes. Results of the meta-analysis showed that teaching children to manipulate the sounds in language helps them learn to read. Across the various conditions of teaching, testing, and participant characteristics, the effect sizes were all significantly greater than chance and ranged from large to small, with the majority in the moderate range. Effects of PA training on reading lasted well beyond the end of training. PA instruction produced positive effects on both word reading and pseudoword reading, indicating that it helps children decode novel words as well as remember how to read familiar words. PA training was effective in boosting reading comprehension, although the effect size was smaller than for word reading. This was not surprising. PA instruction could be expected to benefit children's reading comprehension because of its dependence on effective word reading. However, the NRP had not expected the effect to be as strong, given that the influence is indirect. Other capabilities influence reading comprehension as well, such as children's vocabulary, their world knowledge, and their memory for text. PA instruction helped all types of children improve their reading, including normally developing readers, children at risk for future reading problems, disabled readers, preschoolers, kindergartners, 1st graders, children in 2nd through 6th grades (most of whom were disabled readers), children across various SES levels, and children learning to read in English as well as in other languages.

Does PA instruction assist children in learning to spell? If so, which students are helped?

Yes. Teaching PA was found to help children learn to spell, and its effect lasted well beyond the end of training. Some but not all types of students benefited from PA instruction. It helped kindergartners and 1st graders learn to spell. PA instruction also benefited children at risk for future reading problems and normally developing readers and was effective in boosting spelling skills in low SES as well as mid-to-high SES children. It helped children learning to spell in English as well as children learning in other languages. However, PA instruction was not effective for improving spelling in disabled readers. This is consistent with other research indicating that disabled readers have a difficult time learning to spell.

What properties of instruction make it most effective?

The NRP findings indicate that PA instruction may be most effective when children are taught to manipulate phonemes with letters, when the instruction is explicitly focused on one or two types of phoneme manipulations rather than multiple types, and when children are taught in small groups. Of course, instruction must be suited to students' level of development, with easier PA tasks appropriate for younger children. Teaching with letters is important because this helps children apply their PA skills to reading and writing. Teaching children to blend phonemes with letters helps them decode. Teaching children phonemic segmentation with letters helps them spell. If children have not yet learned letters, it is important to teach them letter shapes, names, and sounds so that they can use letters to acquire PA. PA instruction is more effective when it makes explicit how children are to apply PA skills in reading and writing tasks. PA instruction does not need to consume long periods of time to be effective. In these analyses, programs lasting less than 20 hours were more effective than longer programs. Single sessions lasted 25 minutes on average. Classroom teachers as well as computers can teach PA effectively.

Implications for Reading Instruction

Are the results ready for implementation in the classroom?

Yes. The NRP report includes many ideas that provide guidance to teachers in designing PA instruction and in evaluating existing programs. The NRP has listed references that teachers can locate for additional ideas and guidance. However, there were some important issues not addressed by the research. In implementing PA instruction in the classroom, teachers should bear in mind several serious cautions.

- Teachers should recognize that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping learners understand and use the alphabetic system to read and write. This is why it is important to include letters when teaching children to manipulate phonemes and why it is important to teach children explicitly how to apply PA skills in reading and writing tasks.
- It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness, so PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling already, whereas others may know only a few letters and have no reading skill. Nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA before beginning PA instruction. This will indicate which children need the instruction and which do not, which children need to be taught rudimentary levels of PA (e.g., segmenting initial sounds in words), and which children need more advanced levels involving segmenting or blending with letters.
- PA training does not constitute a complete reading program. Although the present meta-analysis confirms that PA is a key component that can contribute significantly to the effectiveness of

beginning reading and spelling instruction, there is obviously much more that needs to be taught to children to enable them to acquire reading and writing competence. PA instruction is intended only as a critical foundational piece. It helps children grasp how the alphabetic system works in their language and helps children read and spell words in various ways. However, literacy acquisition is a complex process for which there is no single key to success. Teaching phonemic awareness does not ensure that children will learn to read and write. Many other competencies must be taught for this to happen.

- A number of PA instructional programs were found to be effective. The studies assessing these programs are useful in identifying several factors that are important and should be considered in planning classroom instruction or in evaluating published programs that purport to teach PA. In implementing PA instruction in their classrooms, teachers need to evaluate the methods they use against measured success in their own students.
- One factor that is obviously important in any effective classroom program but has not been specifically addressed in the research literature on PA instruction is motivation of the students and of the teachers. It seems self-evident that techniques to develop children's PA in classrooms should be as relevant and exciting as possible so that the instruction engages children's interest and attention in a way that promotes optimal learning. However, research has not specifically focused on this factor. Neither has the research examined the specific techniques that are most engaging for teachers. For example, none of the studies inquired whether teachers liked the programs they were given to teach. It seems self-evident that teachers will be most effective when they are enthusiastic in their teaching and enjoy what they are doing in the classroom. In selecting ways to teach PA in their classrooms, teachers need to take account of motivational aspects of programs for themselves as well as their students.
- Results of the meta-analysis should not be overinterpreted. Although most comparisons in the analysis demonstrated significant mean effect sizes, the NRP cannot infer that every teacher of every child in the studies was successful in promoting the acquisition of PA or its transfer to reading and writing. There was considerable variation within and across individual studies. Likewise, the NRP findings should not be used to dictate any oversimplified prescriptions regarding effective PA instruction, for example, how long PA training should last (e.g., 5 to 18 hours) to be most effective. There are many factors that govern the effectiveness of instruction.
- More is not necessarily better. The NRP findings indicated that PA training was effective regardless of its length. However, effect sizes were largest when training lasted less than 20 hours. This suggests that teachers should make reasoned decisions and remain flexible about the amount of time to devote to this component of their instructional programs. Children will differ in the time they need to acquire PA. The best solution is to pretest for PA skills and adjust the amount of instruction to suit individual and class needs.
- Early PA instruction cannot guarantee later literacy success. The most reasonable conclusion from the findings of the NRP analysis is that adding well-designed PA instruction to a beginning reading program or a remedial reading program is very likely to yield significant dividends in the acquisition of reading and writing skills. Whether the benefits are lasting will likely depend on the comprehensiveness and effectiveness of the entire literacy program that is taught. Additional factors that play a significant role in children's literacy acquisition are detailed in other sections of the NRP report.

Directions for Further Research

Many experiments have been conducted to test whether phonemic awareness instruction helps children learn to read. Results have been sufficiently positive to sustain confidence that this treatment is indeed effective across a variety of child and training conditions. However, there are still some questions needing further attention from researchers.

- Research is needed to identify what teachers need to know and be able to do to teach PA effectively and to integrate this instruction with other elements of beginning reading instruction or instruction directed at older disabled readers.

- Research is needed to study whether small groups are the most effective way to teach phonemic awareness and, if so, the processes and conditions that make this approach especially effective.
- Research is needed to evaluate motivational properties of PA training programs and ways of enhancing motivation and interest if they are lacking. This includes assessing whether approaches appeal to teachers as well as students. It is important to study the factors that influence whether teachers are likely to continue using programs once they are learned.
- Research is needed to determine whether and how PA might be taught more effectively using computers so that transfer to spelling as well as reading is maximized.

PART I: PHONEMIC AWARENESS INSTRUCTION

Report

Introduction

When today's educators discuss the ingredients of effective programs to teach children to read, phonemic awareness (PA) receives much attention. However, not everyone is convinced. In education, particularly in the teaching of reading over the years, the choice of instructional method has been influenced by numerous factors, not only teachers' own frontline experiences about what works, but also politics, economics, and the popular wisdom of the day. Historically, the pendulum has swung back and forth between holistic, meaning-centered approaches and phonics approaches without much hope of resolving disagreements. Meanwhile, substantial scientific evidence has accumulated purporting to shed light on reading acquisition processes and effective instructional approaches (Anderson, Hiebert, Scott, & Wilkerson, 1985; Adams, 1990; Snow, Burns, & Griffin, 1998). Many studies investigating the effectiveness of phonemic awareness instruction have contributed to this body of evidence. Proponents believe that such research holds promise of placing reading instruction on a more solid footing and ending the periodic upheavals and overhauls.

The purpose of this report is to examine the scientific evidence supporting claims about the impact of phonemic awareness instruction on reading development. The National Reading Panel (NRP) sought answers to questions such as the following: Is phonemic awareness instruction effective in helping children learn to read? Under what circumstances and for which children is it most effective? Were studies showing its effectiveness designed to yield scientifically valid findings? What does a careful analysis of the findings reveal? How applicable are these findings to classroom practice?

There were several reasons why the Panel selected phonemic awareness instruction for review and analysis. First, correlational studies have identified phonemic awareness and letter knowledge as the two best school-entry predictors of how well children will learn to read during the first 2 years of instruction (Share, Jorm, Maclean, & Matthews 1984). Such evidence suggests the potential instructional importance of PA training in the development of reading skills. Second, many experimental studies have been conducted to evaluate the effectiveness of PA training in facilitating reading acquisition. Results of these studies claim to be positive and to provide a scientific basis documenting the efficacy of PA training programs. Third, there is currently much interest in PA training programs among teachers, principals, and publishers because of claims about their effectiveness in improving children's ability to learn to read. State adoption committees such as those in Texas and California have prescribed the inclusion of PA training in reading instruction materials approved for use in schools. Thus it is important to determine whether PA training programs live up to these claims and, if so, to identify the circumstances that govern their effectiveness.

In order to evaluate the adequacy and strength of the evidence, the NRP conducted a meta-analysis. The Panel located all of the experimental studies that (1) administered PA training to students, (2) that included control groups, and (3) that measured the impact of training on reading outcomes. The Panel found 52 published studies that met the NRP criteria. The studies varied in many respects. Different types of phonemic awareness skills were taught. The participants ranged from preschoolers to 6th graders and included students at risk for reading problems as well as students classified as reading disabled. The instruction was delivered by classroom teachers in some studies and by researchers or computers in other studies. Children were tutored individually, or they received instruction in small groups, or in larger classroom groups. The meta-analytic procedure allowed the Panel to examine not only whether PA instruction exerted a significant impact on reading across all of these different conditions, but also whether these variations made any difference in the size of the impact.

Assessing and Teaching Phonemic Awareness

To understand how the Panel screened and selected studies that taught PA, it is necessary to clarify what phonemic awareness is and what it is not. Phonemes are the smallest units comprising *spoken* language. English consists of about 41 phonemes. Phonemes combine to form syllables and words. A few words have only one phoneme, such as a or oh. Most words consist of a blend of phonemes, such as go with two phonemes, or check with three phonemes, or stop with four phonemes. In the text below, individual phonemes are represented with IPA (International Phonetic Alphabet) symbols between backslashes (e.g., /g/) to contrast them with letters represented by capitals (e.g., G).

Phonemes are different from graphemes, which are units of written language and represent phonemes in the spellings of words (Venezky, 1970, 1999). Graphemes may consist of one letter, for example, P, T, K, A, N, or multiple letters, CH, SH, TH, -CK, EA, -IGH, each symbolizing one phoneme. Some of the studies reviewed taught children to use letters as aids in distinguishing the separate phonemes in speech. However, the studies the Panel accepted into the database did not go beyond this to teach conventional spelling or text writing.

PA refers to the ability to focus on and manipulate phonemes in spoken words. In the studies reviewed, researchers used the following tasks to assess children's PA or to improve their PA through instruction and practice:

1. Phoneme isolation, which requires recognizing individual sounds in words, for example, "Tell me the first sound in paste" (/p/);
2. Phoneme identity, which requires recognizing the common sound in different words, for example, "Tell me the sound that is the same in bike, boy, and bell" (/b/);
3. Phoneme categorization, which requires recognizing the word with the odd sound in a sequence of three or four words, for example, "Which word does not belong? bus, bun, rug" (rug);
4. Phoneme blending, which requires listening to a sequence of separately spoken sounds and combining them to form a recognizable word, for example, "What word is /s/ /k/ /u/ /l/?" (school);
5. Phoneme segmentation, which requires breaking a word into its sounds by tapping out or counting the sounds, or by pronouncing and positioning a marker for each sound, for example, "How many phonemes in ship?" (3: /š/ /I/ /p/); and
6. Phoneme deletion, which requires recognizing what word remains when a specified phoneme is removed, for example, "What is smile without the /s/?" (mile).

One question of interest in the meta-analysis was whether teaching some forms of PA helped children learn to read better than teaching other forms.

Note that the above list does not include phoneme discrimination, which refers to the ability to recognize whether two spoken words are the same or different, for example, recognizing that tan sounds different from Dan. Phoneme discrimination is simpler than PA because it requires neither conscious awareness of phonemes nor phoneme manipulation. To qualify for analysis, studies had to teach active manipulation of phonemes, not just phoneme discrimination.

Also phoneme awareness is different from phonological awareness, which is a more encompassing term referring to various types of awareness, not only PA but also awareness of larger spoken units such as syllables and rhyming words. Tasks of phonological awareness might require students to generate words that rhyme, to segment sentences into words, to segment polysyllabic words into syllables, or to delete syllables from words (e.g., what is cowboy without cow?). Tasks that require students to manipulate spoken units larger than phonemes are simpler for beginners than tasks requiring phoneme manipulation (Liberman, Shankweiler, Fischer, & Carter, 1974). PA training in the NRP set of studies very often began by teaching children to analyze larger units. For example, Lundberg, Frost, and Petersen (1988) taught children rhyming exercises and how to break sentences into words and words into syllables before they taught children to segment initial phonemes in words. However, if the programs used to teach PA did not progress to the phonemic level, then the study was not included in the NRP data set.

In a few of the studies analyzed by the NRP, instruction was focused on teaching children to manipulate onsets and rimes in words (Fox & Routh, 1984; Lovett, Barron, Forbes, Cuksts, & Steinbach, 1994; Treiman & Baron, 1983; Wilson & Frederickson, 1995). The onset is the single consonant or consonant blend that precedes the vowel, and the rime is the vowel and following consonants, for example, j-ump, st-op, str-ong. Dividing single-syllable words into these units is easier than dividing the words in other places, for example, after the vowel (Treiman, 1985). The NRP included these studies in the set because students were essentially manipulating phonemes when the onset was a single phoneme.

Some forms of PA training in the data set qualified as phonics instruction, which involves teaching students how to use grapheme-phoneme correspondences to decode or spell words. For example, Williams' (1980) ABD program taught students to use graphemes and phonemes to blend words—which is decoding. Ehri and Wilce (1987b) taught students to use graphemes and phonemes to segment words—which is spelling. Also, Wise, King, and Olson (in press) taught both segmenting and blending with letters. What distinguished the NRP studies from the general pool of phonics training studies, however, is that instruction given to treatment students but withheld from controls was limited to grapheme-phoneme manipulation and did not go beyond this to include other activities such as reading decodable text or writing stories.

Contribution of PA in Learning to Read

As mentioned above, PA measured at the beginning of kindergarten is one of the two best predictors of how well children will learn to read. In a study by Share et al. (1984), kindergartners were assessed on many measures when they entered school, including phonemic segmentation, letter name knowledge, memory for sentences, vocabulary, father's occupational status, parental reports of reading to children, TV watching, and many more. These researchers examined which of these measures best predicted how well the children would be reading at the end of kindergarten and at the end of 1st grade. Results showed that PA was the top predictor along with letter knowledge. PA correlated 0.66 with reading achievement scores in kindergarten and 0.62 with scores in 1st grade. Of interest in our analysis was whether PA could be shown to play a causal role in learning to read.

PA is thought to contribute in helping children learn to read because the structure of the English writing system is alphabetic. Moreover, it is not easy to figure out the system. Words have prescribed spellings that consist of graphemes symbolizing phonemes in predictable ways. Being able to distinguish the separate phonemes in pronunciations of words so that they can be linked to graphemes is difficult. This is because spoken language is seamless and there are no breaks in speech signaling where one phoneme ends and the next one begins. Rather phonemes are folded into each other and are coarticulated. Discovering phonemic units is helped greatly by explicit instruction in how the system works. This is underscored by research revealing that people who have not learned to read and write have great trouble performing phonemic awareness tasks (Morais, Bertelson, Cary, & Alegria, 1987). Likewise people who have learned to read in a script that is not graphophonemic, such as Chinese, have difficulty segmenting speech into phonemes (Mann, 1987; Read, Zhang, Nie, & Ding, 1987). For these reasons, it was expected that the impact of PA training on literacy would be strongest in tasks assessing children's ability to read and spell words.

Research on word reading processes has distinguished several ways to read words (Ehri, 1991, 1994). The process of decoding words never read before involves transforming graphemes into phonemes and then blending the phonemes to form words with recognizable meanings. The PA skill centrally involved in decoding is blending. To assess decoding skill, researchers often test children's ability to read pseudowords such as blig or nef.

A second way to read unfamiliar words is by analogy to known words (Gaskins, Downer, Anderson, Cunningham, Gaskins, Schommer, & the Teachers of Benchmark School, 1988; Glushko, 1979; Goswami, 1986; Marsh, Freidman, Welch, & Desberg, 1981). A common basis for analogizing is recognizing that the rime segment of an unfamiliar word is identical to that of a familiar word, and then blending the known rime with the new onset, for example, reading brick by recognizing that -ick is contained in the known word kick. Reading by analogy is thought to require the PA skills of onset-rime segmentation and blending.

Another way to read words is from memory, sometimes called sight word reading. This requires prior experience reading the words and retaining information about them in memory. In order for individual words to be represented in memory, beginning readers are thought to form connections between graphemes and phonemes in the word. These connections bond spellings to their pronunciations in memory (Ehri, 1992; Ehri & Wilce, 1987a; Rack, Hulme, Snowberg, & Wightman, 1994; Reitsma, 1983). The PA skill thought to be important for developing word memory is being able to segment pronunciations into phonemes that link to graphemes. Formulation of this concept led to the expectation PA training would benefit children's word reading, particularly when they received practice learning to read the words.

The processes involved in writing words, either by generating approximate spellings of the words or by retrieving correct spellings from memory, require phonemic segmentation skill (Griffith, 1991). Phonemic segmentation is required for spellers to select letters to represent the phonemes. Phonemic segmentation is required to help children retain correct spellings in memory by connecting graphemes to phonemes. In the analysis it was expected that PA training would benefit children's ability to spell.

Various kinds of word reading outcomes were assessed across the studies the Panel reviewed. The simplest task given to preschoolers required them to look at a word (sat) and decide whether it says sat or mat (Byrne & Fielding-Barnsley, 1991). Studies with older children gave them lists of words to read either from standardized tests or experimenter-devised tests. Also, word learning tasks were used. For example, kindergartners first reviewed four letter-sound relations and then practiced learning to read five words over several trials, am, at, mat, sat, Sam (O'Connor, Jenkins, & Slocum, 1995). Also, pseudoword reading tasks were used in which children read nonwords such as feem, hote, cliss. Spelling tasks were included as well. Younger children were given credit for inventing phonetically plausible spellings of words while older children were scored for producing correct spellings.

Some of the studies in the NRP database measured reading comprehension as well as word reading. In order to comprehend a text, readers must be able to read most of the words. However, other capabilities influence reading comprehension as well, such as readers' vocabulary, their world knowledge, and their memory for text. It was expected that PA training would benefit children's reading comprehension because of its dependence on effective word reading. However, the degree of influence was expected to be less than that observed with word reading because the influence is indirect.

Design Features of Phonemic Awareness Training Studies

Many correlational studies have reported strong relationships between phonemic awareness and learning to read (for reviews, see Blachman, in press; Ehri, 1979; Stahl & Murray, 1994; Wagner & Torgesen, 1987). In correlational studies, researchers measure children's ability to manipulate phonemes and also their reading ability. Typical findings show that students who have superior phonemic awareness are better readers than students with low PA. However, such findings are insufficient to show that PA was the underlying cause enabling some students to read better than others. This is because the finding does not rule out other causal explanations for the relationship. Perhaps the correlation was observed because cause operated in the reverse direction; that is, learning to read improved students' PA. Or perhaps a third factor operated as an underlying cause boosting both PA and reading, for example, vocabulary size, memory, or general intelligence.

In order to show that PA operates as a direct cause in helping children learn to read, the NRP needed to assess evidence from experimental studies with treatment and control groups. A well-designed experiment that provides strong evidence for cause should include the following steps:

1. Pretesting should be given to students before they receive any training. Pretests verify that children have not already acquired PA and hence can profit from training. Pretest performance can be compared to posttest performance on PA, reading, and spelling tasks to evaluate gains resulting from PA training. Also, pretests indicate whether

treatment and control groups were equivalent prior to training. If not, pretests can be use to equate the groups statistically when effects of training are evaluated on outcome measures.

2. The group receiving PA training should be compared to a control group that is equivalent in all respects except for receiving the PA training. Control groups may receive another type of training involving equal time but no PA instruction, or control groups may receive no special training beyond that provided in the students' classrooms at school. The use of an alternative-treatment control group is considered preferable to a no-treatment control group because the former rules out the Hawthorne effect as the explanation for any outcome differences favoring the experimental group. The Hawthorne effect occurs when a treatment group outperforms a no-treatment control group because the treated group received special attention and as a result was more motivated to perform.
3. Random assignment should be used to place students in treatment and control groups. Random assignment makes it likely that treatment and control groups do not differ systematically in any way that would explain outcome differences following training. In other words, this step helps to establish that the treatment, rather than some other factor, was the cause of any improvement in reading outcomes.
4. Posttests should be given to students following training. Posttests to assess PA verify that training worked, that the PA-trained group made greater gains than the control group. Posttests to assess reading and spelling show that PA training transferred and improved students' reading and spelling performance.
5. Followup posttests should assess the long-term effects of PA training on students' progress in reading and spelling. Between the end of training and the followup tests, both experimental and control students receive regular instruction at school but no further specialized training in PA.

Although these features characterize a well-designed experiment, there were studies in the NRP database that lacked some of these features. Because of this, the relationship between design features and outcomes was assessed. Studies varied in whether they compared performance of the PA-trained groups to performance of treated control groups or untreated control groups. If Hawthorne effects have influenced comparisons, one would expect bigger effects when PA treatment groups are compared to untreated control groups than when compared to treated control groups. However, Bus and van Ijzendoorn (1999) in their meta-analysis reported the reverse, finding bigger effects in comparisons between PA treatment groups and control groups receiving an alternative treatment. The Panel attempted replication of their findings with the NRP data set.

The Panel also assessed whether PA training affected outcomes in three types of designs: (1) in true experiments where students were randomly assigned to treatment and control groups; (2) in quasi-experiments where students were members of pre-existing groups which were not randomly assigned to treatment and control conditions; and (3) in studies where students from treatment and control groups were matched. Although random assignment is preferable, researchers may be limited to a quasi-experimental design when they evaluate PA programs in schools where classrooms already exist or when they employ as trainers teachers who are already familiar with a program and teach it to their students. The procedure of matching children on the basis of pretest scores is done to minimize any pretreatment differences between the groups being compared. In the NRP analysis, the effects of PA training separately for the three types of studies were examined.

In a recent critique of PA training studies, Troia (1999) identified several design flaws and applied these criteria to rate PA training studies for their lack of methodological rigor. To evaluate the impact of these flaws on outcomes, the Panel examined the relationship between Troia's assessments of the PA studies and the effects reported in these studies. The purpose of this analysis was to rule out the possibility that claims about PA training effects are supported mainly by poorly designed studies.

Other Features of PA Training Studies

Studies in our data set varied in the types of students who received PA training. The NRP wanted to know whether certain types of students benefited more than other types. Studies varied in the grade level of their participants and ranged from preschool to 6th grade. Studies varied in whether their students showed any signs of having reading problems. Three types of readers were distinguished across the studies. Some focused on children at risk for developing reading difficulties in the future. These were children below 2nd grade. Being at risk was defined as having low PA or low reading in 83% of the cases. Low socioeconomic status (SES) characterized only 27% of the cases. Some studies focused on children who had already fallen behind classmates in their reading, referred to as disabled readers. These were children in 1st grade and above. The remaining studies sampled children who were judged to be making normal progress in learning to read. This judgment was based on the fact that the children were not identified as having any reading problems.

One common finding reported in many correlational studies is that children who are or will become disabled readers have poor phonemic awareness, substantially below that expected of students at their reading levels (Bradley & Bryant, 1983; Bruck, 1992; Fawcett & Nicholson, 1995). Researchers have suggested that this deficiency underlies and explains their difficulty in learning to read. In the NRP analysis, the Panel examined whether PA training was effective in teaching PA to at-risk and disabled readers and whether this improved their reading and spelling performance, thus providing evidence for a causal connection.

Studies varied in how the PA training was delivered. In some studies, researchers or their specially trained assistants taught children to manipulate phonemes. In other studies, classroom teachers were the trainers. In a few studies, training was presented primarily by computers. Because classroom teachers are the purveyors of reading instruction for most children, it is important to determine whether they can teach PA effectively. If training requires specially trained personnel, then PA instruction should not be imposed on classroom teachers. In the NRP analyses, the effects of PA training were examined separately for teachers, for computers, and for researchers.

There is substantial evidence that one-to-one tutoring is the most effective form of instruction (Bloom, 1984; Cohen, Kulik, J., & Kulik, C., 1982; Glass, Cahen, Smith, & Filby, 1982; Pinnel, Lyons, DeFord, Byrk, & Seltzer, 1994; Wasik & Slavin, 1993). However, Bus and van Ijzendoorn (1999), in their meta-analysis of PA training studies, found that teaching PA to small groups of children produced a bigger impact on outcomes than teaching students individually or in classrooms. The aim was to attempt replication of this finding with the NRP data set that included more studies than those in the previous meta-analysis.

It is common wisdom that greater time spent training students yields superior learning. However, instructional time in schools is very limited because of the many subjects and skills that must be taught. The studies in the NRP data set varied in the length of time spent teaching PA to students. To address the question of how much time might be sufficient for teaching PA, the relationship between training time and effects on learning was examined.

The NRP database included PA training studies conducted not only in English but also in other languages, such as Norwegian, Finnish, Swedish, Danish, Spanish, Hebrew, Dutch, and German. In most of these languages, the grapheme-phoneme connections are more transparent than in English. Of interest was whether PA training might exert a larger impact in English because it is harder for beginning readers to discover the graphophonemic system in English than in other languages.

Methodology

Database

An electronic search of two databases, ERIC and PsycINFO, was conducted. Six terms involving phonemic awareness were crossed with 15 terms related to reading performance. The PA terms were: phonemic awareness, phonological awareness, spelling, blending, learning to spell, and invented spelling. The reading terms were: reading, reading ability, reading achievement, reading comprehension, reading development, reading disabilities, reading skills, remedial reading, beginning reading, beginning reading instruction, reading acquisition, word identification, word reading, oral reading, and miscues. The search was limited to articles appearing in journals written in English, but no limit was placed on the year of publication. Using this procedure, the Panel located 637 articles through ERIC, and 1,325 articles through PsycINFO. Abstracts were printed and screened. In addition, the Panel hand-searched and screened references cited in the studies located by the electronic search and in several review papers (Apthorp, 1998; Blachman, in press; Bus & van Ijzendoorn, 1999; Stahl & Murray, 1994; Troia, 1999; Wagner, 1988).

To qualify for the analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group or a multiple baseline method.
2. Studies had to appear in a refereed journal.
3. Studies had to test the hypothesis that training in phonemic awareness improves reading performance over alternative forms of training or no training.
4. Studies had to provide training in phonemic awareness that was not confounded with other instructional methods or activities.
5. Studies had to report statistics permitting the calculation or estimation of effect sizes.

From the various lists of references, the Panel identified and located 78 articles that appeared to meet our criteria. Upon closer inspection, 26 articles did not match all criteria: 5 lacked sufficient information to determine effect size; 5 lacked an adequate control group; 12 did not assess reading as an outcome; and 4 lacked appropriate phonemic awareness training. The final set of studies meeting our criteria numbered 52 (see Appendix A).

The primary statistic used in the Panel's analysis of performance on outcome measures was effect size, indicating the extent to which performance of the treatment group exceeded performance of the control group, with the difference expressed in standard deviation units. The formula used to calculate raw effect sizes for each treatment-control comparison consisted of the mean of the treatment group minus the mean of the control group divided by a pooled standard deviation.

From the 52 studies, 96 cases comparing individual treatment and control groups were derived. Because some of the studies included more than one treatment or control group, the cases included comparisons utilizing the same group more than once. There were seven treatment groups appearing twice because they were compared to two different control groups. There were 16 control groups appearing twice because they were compared to 2 different treatment groups. There was one control group appearing three times because it was compared to three treatment groups. In sum, there were 47 independent comparisons and 49 comparisons having a group that overlapped with one or at most two other comparisons. Although this meant that effect sizes were not completely independent across cases, the Panel preferred this alternative to combining treatment and control groups within studies because it was important not to obscure important moderator variables of interest. For example, Davidson and Jenkins (1994) studied three treatment groups, one taught to blend, one taught to segment, and one taught to both to segment and blend. They compared the performance of each treatment to the same control group. The Panel wanted to retain these as separate comparisons in our analysis, so the same control group was allowed to recur in three comparisons.

A few studies in the NRP database included treatment or control groups that were not deemed appropriate for analysis. One reason was that the treatment groups provided not only phonemic awareness training but also reading or writing training that was not provided to control groups, thus confounding PA training with reading and writing training. The following describes which treatment or control groups were eliminated from the analysis and why: a treatment group given decoding training and word reading (Barker & Torgesen, 1995); a treatment group given a reading and writing program (Brennan & Ireson, 1997); a treatment group taught to manipulate syllables rather than phonemes (Sanchez & Rueda, 1991); a treatment group taught semantic categorization with written words (Defior & Tudela, 1994); treatment groups in which the teacher-trainers failed to spend the time prescribed for training (Olofsson & Lundberg, 1983); treatment groups in which children not only analyzed phonemes but also read words in sentences and stories, unlike children in the control groups who only listened to stories or remained in their classrooms (Solity, 1996; Weiner, 1994); a control group lacking not only PA training but also the Reading Recovery® instruction given the treatment group (Iversen & Tunmer, 1993); and a control group that did not control for all of the non-PA elements of training (Lovett et al., 1994; Vellutino & Scanlon, 1987). These treatment or control groups were not included in the database.

The studies in the NRP database were coded for many characteristics that the Panel felt were important to include as moderator variables in the meta-analyses. These characteristics are listed in Table 1 (Appendix B). Various properties of phonemic awareness training were coded. Training programs varied in whether they focused on specific PA manipulations. Single-focus studies taught blending, categorization, identity, segmention, or onset-rime only. Double-focus studies involved combinations of blending, segmenting, deletion, or categorization. Global treatments taught three or more PA skills. Programs that only taught onset-rime manipulation were coded as onset-rime training, even though the training might have involved blending and segmenting (e.g., Fox & Routh, 1976). Training varied in whether children were taught to manipulate phonemes using letters or whether attention was limited to phonemes in speech. Training that had children manipulate blank markers was coded as a nonletter treatment.

The training unit varied across studies. Students were tutored individually in some studies and in either small groups or whole classrooms in other studies. The size of the small groups varied from two to seven students. The identity of trainers varied across studies. The Panel compared classroom teachers to others who were mostly researchers or trained assistants. Credentialed teachers who conducted the training but were not the students' classroom teacher were coded as others. In a few studies, PA training was provided mainly by computers. The Panel compared this training to training provided by noncomputers (all others). The length of training varied from 1 to 75 hours. Comparisons were conducted by dividing training time into four blocks.

Characteristics of children receiving the training were coded. Children were grouped into four categories to reflect their grade levels: preschool, kindergarten, 1st grade, and 2nd through 6th grades. Also children were grouped by reading ability. At-risk children were those judged by authors of the studies to be at risk for developing reading problems. In the majority of cases (77%), this was indicated by poor performance on PA tasks. Other indicators used in a few studies were low reading, low SES, developmental or language delays, or cognitive disabilities. Only 27% of the cases were low SES, while 37% were middle-to-high SES. These children were all below 2nd grade.

Children who had already developed reading problems were coded as disabled readers. All but three cases involved children between 2nd and 6th grade levels. The three cases involved 1st graders who qualified for Reading Recovery® programs (Hatcher, Hulme, & Ellis, 1994; Iversen & Tunmer, 1993). Being reading disabled meant reading below grade level despite at least average cognitive ability in most studies. In one study, the school's definition of learning disabled was used (Williams, 1980). In one study, students were not only reading disabled but also had neurological impairment and language learning problems (Lovett et al., 1994).

Samples of children not reported as being at risk or reading disabled were coded as normally progressing readers. These studies included children selected not to have reading problems as well as children selected without regard to reading ability. The socioeconomic level of children was coded into two categories, low SES or middle-to-high SES, based on assertions by authors. The language spoken by children and used to teach PA was coded as English or non-English. Non-English languages included Dutch, Finnish, German, Hebrew, Norwegian, Spanish, and Swedish.

Some features of the methodology used in the experiments were coded. Children were assigned to treatment and control groups in one of three ways. They were randomly assigned. Or they were members of intact groups that were not randomly assigned to conditions, referred to by researchers as nonequivalent groups. In some studies two classrooms were assigned randomly, one to the treatment and one to the control condition. These cases were categorized as nonequivalent groups. In other studies, several classrooms were assigned randomly to treatment and control conditions. These cases were categorized as random assignment. The third way of assigning children to conditions involved matching children on the basis of similar test scores. Typically, members of a match are randomly assigned, one to the treatment group and one to the control group. However, in some studies, this step was not stated explicitly; so, it is impossible to be sure that random assignment was always used.

The Panel coded studies to reflect whether fidelity to treatment was checked, that is, whether researchers observed trainers to make sure they adhered to treatment procedures. In addition, comparisons were coded for the type of control group, that is, whether or not control students received a special alternative treatment or remained untreated. The number of students participating in the comparison was coded to reflect sample size. The numbers were grouped into four blocks to distinguish sample sizes ranging from small to large.

To evaluate the relationship between the methodological quality of studies and the effect sizes found, the Panel adopted the five methodological criteria applied by Troia (1999) in his critique of the internal and external validity of PA training studies. Internal validity refers to the authenticity of cause-and-effect relationships in a study, that is, whether the treatment caused the outcome observed, or whether other variables could have impacted the outcome. External validity refers to the generalizability of the findings, that is, whether or not the results of a study can be applied to other persons in other settings at other times. To evaluate the internal and external validity of studies, Troia used four summary measures: percentage of internal validity criteria met by the studies, number of critical flaws challenging a study's internal validity (e.g., no random assignment, no alternative treatment given to the control group, no assessment of trainer fidelity to treatment), percentage of external validity criteria met, and number of critical flaws challenging a study's external validity (e.g., insufficient information about the sample of participants or about how disability was defined and assessed). Troia evaluated 28 of the studies included in the NRP database. The Panel applied his ratings and rankings to the 56 cases derived from these studies. The Panel did this without checking Troia's evaluations for accuracy; so, any incorrect codings of the studies arise from Troia's procedures, not from the Panel's.

One final characteristic of the NRP studies was coded and analyzed, the year of publication. Years were cast into four blocks. Other characteristics of the studies were coded as well but were not analyzed either because there was little interest or because there was an insufficient number of cases to support a meaningful analysis.

Four individuals coded the studies and entered values into the SPSS database. The reliability of moderator-variable codes was checked by comparing codes in the database to codes generated by one of the coders who re-coded 14 of the articles (15% of the cases). The percentage of agreement of the codes was 94%. All of the means, standard deviations, and sample sizes that were entered into the database were verified at least twice for accuracy.

There were three outcomes of primary interest: phonemic awareness, reading, and spelling performance. Some studies included multiple tasks measuring these outcomes. These measures were combined by calculating raw effect sizes (g) for individual tasks and then averaging the effect sizes across tasks. The composite measure for reading included many different types and measures of reading. For example, word reading, pseudoword reading, reading comprehension, reading speed, time to reach a criterion of learning, and miscues were included. The phonemic awareness composite included only those measures that required manipulating phoneme-size units, not larger syllabic units. The types of manipulations in the composite included segmentation, blending, reversing, deletion, identity, and categorization. The spelling composite included measures of the quality of invented spellings as well as correct spellings of words and pseudowords.

The Panel also examined more specific outcome measures that included various types of phonemic awareness, reading, spelling, and math. The specific measures are listed in Table 1. Also of interest was a comparison of effect sizes on outcomes measured immediately after training to outcomes assessing long-term learning. Delayed posttests were administered from 2 to 36 months following training.

Meta-Analysis

Most of the studies in the NRP database reported treatment and control group means and standard deviations that were used to calculate effect sizes. However, there were 14 studies that lacked sufficient information. DSTAT was employed (Johnson, 1989) to estimate these effects, usually from F- or t- or MSE values, or the information was obtained from authors.

The analysis of effect sizes across studies was conducted by giving more weight to effect sizes that were based on larger samples of participants. However, the following studies administered training to groups of students and hence used groups rather than individual students as the unit of analysis in their statistics: Byrne & Fielding-Barnsley, (1991); Castle, Riach, & Nicholson, (1994); O'Connor, Jenkins, & Slocum, (1995); Torgesen, Morgan, & Davis, (1992); Williams, (1980) (Experiment 2). Using the number of groups as the value of n in the weighting procedure for these studies had the effect of underrepresenting their effect sizes. To address this problem, the Panel used n's for the unit of analysis to convert raw effect sizes (g) to corrected effect sizes (d) in each case. Then, when composite effect sizes were calculated across cases, the individual effect sizes (d) were weighted by the number of students in the sample, not by the unit of analysis, thus ensuring that no cases were underrepresented.

The DSTAT statistical package (Johnson, 1989) was employed to determine effect sizes and to test the influence of moderator variables on effect sizes. Each moderator variable had at least two levels. The Panel tested whether the mean weighted effect size (d) at each level was significantly greater than zero at $p < 0.05$, whether the individual effect sizes at each level were homogeneous ($p < 0.05$), and whether effect sizes differed significantly at different levels of the moderator variables ($p < 0.05$).

Consistency With the Methodology of the National Reading Panel

The NRP review methodology (NRP Progress Report, February 1999) was used in the search and analysis of the studies. Specifically, studies that were not published in peer-reviewed journals were excluded. All of the studies in the database employed experimental or quasi-experimental designs. The studies were coded for most of the specified categories. Categories left uncoded were those where information was rarely provided (e.g., setting [urban, rural, suburban], cost factors associated with training).

The Panel determined that a meaningful meta-analysis could be conducted on the data. The coding of moderator variables and the means and standard deviations that were used to calculate effect sizes were verified by checking all of them at least twice. Intercoder reliability was conducted on the moderator variables and agreement exceeded the prescribed level of 90%. The data analysis followed the procedures specified.

Results

Were Effect Sizes Greater Than Zero?

The statistic used to assess the effectiveness of PA training on outcome measures was effect size that measures how much the mean of the PA-trained group exceeded the mean of the control group in standard deviation units. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of training. An effect size of 0 indicates that treatment and control group means were identical, revealing that training had no effect. To judge the strength of an effect size, values suggested by Cohen (1988) are commonly used. An effect size of 0.20 is considered small; a moderate effect size is 0.50; an effect size of 0.80 or above is large.

Mean effect sizes obtained for outcome measures and levels of the moderator variables are reported in Appendix C—Table 2 for phonemic awareness, Table 3 for reading, and Table 4 for spelling. Effect sizes were tested statistically to determine whether each was significantly greater than zero, indicating that superior performance of PA trained groups over control groups was not likely a result of chance at $p < 0.05$. Inspection across Tables 2 and 3 in Appendix C reveals that all of the effect sizes involving phonemic awareness and reading outcomes were significantly greater than zero. This indicates that training was effective in teaching phonemic awareness and in facilitating transfer to reading across all of the conditions and characteristics considered.

Inspection of spelling outcomes in Table 4 reveals that all but three effect sizes were significantly greater than zero. This indicates that, across most of the conditions and characteristics considered, phonemic awareness training transferred and improved spelling skills more than alternative forms of training or no training. Effect sizes for spelling outcomes were insignificant when computers were used in the training, and when the students trained were disabled readers or children in 2nd grade and above. As documented below, the absence of significant effects on spelling outcomes in the latter cases arose primarily because disabled readers' spelling benefited little from PA training, and these readers were overrepresented in these categories (i.e., 2nd through 6th graders, receiving PA instruction on computers).

Some of the studies evaluated the effects of PA training on an outcome not expected to be affected (e.g., mathematics). Tests to assess math were administered following training in 12 comparisons and following some delay in three comparisons. Results in Table 3 show that the effect size was nonsignificant and close to zero ($\underline{d} = 0.03$). This indicates that the effects of PA training did not influence all outcomes but rather were limited to outcomes related to literacy. These findings argue against the operation of any halo/Hawthorne effect explaining the positive effect sizes.

In sum, these findings led the Panel to conclude with much confidence that phonemic awareness training is more effective than alternative forms of training or no training in helping children acquire phonemic awareness and in facilitating transfer of PA skills to reading and spelling. PA training improves children's reading performance in various types of tasks, including word reading, pseudoword reading, and reading comprehension. Benefits are evident on standardized tests as well as experimenter-designed tests of reading and spelling. Improvement in reading and spelling is not short-lived but lasts beyond the immediate training period.

PA training improves reading performance in preschoolers and elementary students, and in normally progressing children, as well as in older disabled readers and younger children at risk for reading difficulties. PA training improves spelling performance in kindergartners, 1st graders, and at-risk students, but not in older disabled readers. PA training boosts reading and spelling in both English and non-English languages, and among low SES as well as middle-to-high SES children. Many types of PA training programs are effective for improving reading and spelling, including those that teach one or multiple types of phonemic awareness, those that incorporate letters into training, and those that limit phoneme manipulation to speech. Not only researchers but also classroom teachers and computers can deliver PA instruction effectively. Instruction can be conducted successfully with individuals as well as small groups and whole classrooms. Training does not have to be lengthy to be effective.

Were Effect Sizes Homogeneous?

In addition to determining whether mean effect sizes were significant, the Panel also tested whether the set of effect sizes was sufficiently homogeneous to render the mean effect size representative of that set. A homogeneity analysis calculates how probable it is that the variance exhibited among the effect sizes would be observed if only sampling error was making them different (Cooper, 1998). The 95% confidence intervals for effect sizes presented in Tables 2 to 4 reveal how variable they were. When the pool of effect sizes is not homogeneous, the next step is to examine whether moderator variables reduce the variability among effect sizes to create homogeneity, indicating their power to explain the variance.

At the top of Tables 2, 3, and 4 in Appendix C, it is apparent that on the immediate outcome measures of PA, reading, and spelling, effect sizes were not homogeneous, as indicated by "No" in the homogeneity column. Effect sizes involving followup measures of PA and spelling outcomes were homogeneous, but followup reading effect sizes were not. Thus, there is reason to examine moderator variables that may explain effects on immediate outcomes and on followup tests involving reading outcomes.

Did Moderator Variables Influence Effect Sizes?

Studies varied in many respects as indicated in Table 1 (Appendix B). The Panel examined whether these moderator variables enhanced or limited the effectiveness of PA training for teaching PA and for facilitating transfer to reading and spelling. It is important to recognize the limitations of this type of analysis and the tentative nature of any conclusions that are drawn. Findings involving the impact of moderator variables on effect sizes cannot support strong claims about causality. Moderator findings are no more than correlational. The biggest source of uncertainty is whether there is a hidden variable that is confounded with the variable in focus and is the true cause of the difference; thus, the conclusions drawn should be regarded as tentative and suggestive rather than the final word.

Another caution to keep in mind in interpreting findings involving moderator variables is that the same 96 cases in the database do not contribute to the calculation of all effect sizes. Rather the set of cases changes across moderator variables, either because some of the studies lacked the information to be coded, they did not assess the outcome in interest, or they did not include a measure of the outcome at that test point. Any instability in the pattern of findings may arise from this source, particularly when only a few cases contribute.

Outcome Measures

The immediate goal of phonemic awareness training across these studies was to improve children's phonemic awareness. From Table 2, it is apparent that the effect size after training was large ($\underline{d}$ = 0.86), and it did not decline significantly at the followup test ($\underline{d}$ = 0.73). Thus, PA training taught phonemic awareness very effectively, and students retained their skill after training ended. Comparison of specific PA skills acquired during training indicated that effects were larger for segmentation and deletion outcomes than for blending. Perhaps blending was harder to teach, or perhaps it was easier for controls to pick up without instruction.

The strong gains in PA were observed to transfer to reading and spelling, and effects persisted through the second followup test. As evident in Table 3, reading-outcome effect sizes were moderate, and the effect size after training ($\underline{d}$ = 0.53) was equivalent to that at the first followup test ($\underline{d}$ = 0.45). A significant effect size was still present but significantly smaller at the second followup test ($\underline{d}$ = 0.23). Table 4 shows that spelling outcomes were boosted by PA training. The effect size following training ($\underline{d}$ = 0.59) was moderate and significantly greater than the effect sizes at the two delayed posttests ($\underline{d}$ = 0.37 and 0.20) that did not differ.

PA training benefited children's reading and spelling performance not only on experimenter-devised (E) tests but also on standardized (S) tests, although the effect size was significantly larger with experimenter tests ($\underline{d}$ = 0.61 E vs. 0.33 S for reading; $\underline{d}$ = 0.75 E vs. 0.41 S for spelling). This is perhaps not surprising. Standardized tests are designed to assess reading and spelling across a wide range of ability levels and hence are less sensitive to differences at any one level in the range. Also, experimenter tests may be more sensitive because often they are tailored to detect the phonemes and graphemes that were taught.

Some studies assessed reading performance with pseudowords in order to measure children's ability to decode unfamiliar words. From Table 3, it is apparent that PA training benefited decoding skill. Effects were moderate and equivalent on both experimenter-devised tests ($\underline{d}$ = 0.56) and standardized tests ($\underline{d}$ = 0.49).

The effect of PA training on reading comprehension was assessed in 18 cases. From Table 3, it is apparent that training boosted reading comprehension significantly ($\underline{d}$ = 0.32), although the effect size was smaller than for word reading. This is not surprising. PA training would be expected to influence comprehension primarily through its impact on word reading. The task of reading, understanding, and remembering information in the text involves multiple processes. Not only must students read the words, but also they must do so rapidly and accurately and must construct meaning across the words and sentences. These other processing demands could be expected to dilute the influence of PA training.

Properties of PA Training

Studies varied in whether one skill, two skills, or multiple skills were taught. These skills consisted mainly of teaching children to identify or categorize phonemes, or to blend, segment, or delete phonemes, or to manipulate onset-rime units. From PA outcomes in Table 2, it is apparent that focusing instruction on one or two skills was significantly more effective for teaching phonemic awareness than focusing on multiple skills ($\underline{d}$ = 1.16 for one vs. $\underline{d}$ = 1.03 for two vs. $\underline{d}$ = 0.70 for multiple). One explanation for lower effect sizes is that children who were taught many different ways to manipulate phonemes may have become confused about which manipulation to apply when the various kinds of PA were assessed after training. Another possibility is that insufficient time was spent on any one type of PA to teach it well in the multiple condition. A third possibility is that multiple skills instruction involved teaching higher level PA skills mainly to older children having difficulty acquiring PA.

The Panel examined whether focused training in PA produced greater transfer to reading than multiple-skill training. From reading outcomes in Table 3, it is apparent that transfer was twice as great when PA training focused on one (d = 0.71) or two (d = 0.79) PA skills than when a multitude of skills were taught (d = 0.27). The advantage of focused over multiple-skill training for reading persisted at the followup test, especially for the two-skill focus that produced significantly larger effects than the one-skill focus. This indicates that teaching two PA skills to children has greater long-term benefit for reading than teaching only one PA skill or teaching a global array of skills.

As evident in Table 4, spelling effect sizes for focused and multiple skills instruction showed the same pattern. In fact, effects for the one-skill condition (d = 0.74) and the two-skill condition (d = 0.87) were over three times as large as the effect size for the multiple condition (d = 0.23). These findings suggest that focused PA instruction may benefit spelling more than multiple skill instruction does. However, it is likely that the lower effect size in the multiple condition arose because disabled readers dominated this category and PA instruction did not improve their spelling (see below).

Various types of phoneme manipulations might be taught. However, two types, blending and segmenting, are thought to be directly involved in reading and spelling processes. Blending phonemes helps children to decode unfamiliar words. Segmenting words into phonemes helps children to spell unfamiliar words and also to retain spellings in memory. A number of studies examined PA training that taught children to blend and segment phonemes. To assess its value, the Panel compared the effect size for this treatment to the effect size for the multiple (3 or more skills) treatment. As evident in Table 2 reporting PA outcomes, neither form was more effective than the other for teaching PA. However, as evident in Table 3 for reading outcomes, teaching students to blend and segment benefited their reading much more (d = 0.67) than did a multiple-skills approach (d = 27). As shown in Table 4, the blending and segmenting treatment also produced a larger effect on spelling performance (d = 0.79) than did the multiple skill treatment (d = 0.23), but very likely this resulted from disabled readers' dominating the multiple treatment condition (see below). From these findings, the Panel concludes that blend-and-segment training benefited children's reading more than multiple skills training did.

Also of interest was whether some types of single phoneme manipulation activities, for example, blending, segmenting, or categorizing, were more effective than other types. However, in examining the database, there were too few instances of each type to permit comparison; so, this question was not addressed in the Panel's analysis.

Studies in the database differed in whether or not children were taught to manipulate phonemes using letters during training. For example, some children learned to segment words into phonemes by selecting plastic letters for the sounds they spoke, whereas other children only spoke the sounds or they represented the sounds with unmarked tokens. Of interest was whether letters might improve children's learning because they provide concrete, lasting symbols for sounds that are short-lived and hard to grasp. From PA outcomes in Table 2, it is apparent that children trained with letters did not acquire stronger PA (d = 0.89) than children trained without letters (d = 0.82). The absence of a difference may have occurred, however, because almost all comparisons involving disabled readers fell in the letter use category, and disabled readers exhibited smaller effect sizes than nondisabled readers on PA outcomes (see Table 2). As described below, when effects of letter use were examined after disabled readers were removed from the database, a significant advantage of letter use was detected. From these findings, the Panel concludes that teaching PA with letters is more effective in helping nondisabled readers acquire phonemic awareness than teaching PA without letters.

It was expected that teaching PA with letters would facilitate greater transfer to reading and spelling than teaching PA without letters. This is because reading and spelling processes require knowing how phonemes are linked to letters. From reading outcomes in Table 3, it can be seen that teaching children to manipulate phonemes with letters created effect sizes almost twice as large as teaching children without letters (d = 0.67

vs. 0.38). The same pattern persisted at the followup test as well (d = 0.59 vs.0.36). Likewise, letters benefited spelling more than no letters, with the effect size almost twice as great (d = 0.61 vs. 0.34). These findings reveal that PA training makes a stronger contribution to reading and spelling performance when the training includes teaching children to manipulate phonemes with letters than when training is limited to speech.

Studies varied in whether PA training was provided to individual students or small groups or classrooms of students. From PA outcomes in Table 2, it is evident that the most effective way to teach PA was in small groups. The effect size produced by small groups was very large (d = 1.38), over twice the size of effects for individuals (d = 0.60) and classrooms (d = 0.67). This was surprising given that it is easier to tailor instruction and corrective feedback when students are taught individually, and it was expected that this advantage would make individual instruction more effective. Explanations for the effectiveness of training in groups promoting the acquisition of PA may involve enhanced attention, social motivation to achieve, or observational learning opportunities.

The superior PA skills acquired by children taught in small groups transferred and boosted their reading and spelling performance as well. Effect sizes on reading outcomes for small groups were d = 0.81 on the immediate posttest and d = 0.83 on the followup posttest. In contrast, effect sizes for children taught individually or in classrooms ranged from d = 0.30 to 0.45 on the immediate and delayed posttests. On spelling outcomes, small group instruction produced a larger effect size than individual instruction did, but the small group effect size did not differ from the classroom effect size (see Table 4).

The possibility that small group effect sizes might be inflated for statistical reasons was considered. Studies that treated groups as the unit of analysis in statistical comparisons may have exhibited larger effect sizes than studies using individuals as the unit of analysis because the standard deviations of group means are smaller than the standard deviations of individual scores. However, there were only five studies that used groups as the statistical unit of analysis, and these contributed only seven cases (15%) to the total of 45 cases in which children were trained in small groups. The small number of instances serves to rule out this explanation for the larger effect sizes associated with small group training.

The length of time allocated for PA training varied from 1 hour to 75 hours across studies. Cases were grouped into four time blocks to determine whether there was an optimum length of time for teaching PA. From phonemic awareness outcomes in Table 2, it is evident that effect sizes were significantly larger for the two middle time periods lasting from 5 to 9.3 hours (d = 1.37) and from 10 to 18 hours (d = 1.14). Periods that were either shorter or longer than this were less effective for teaching PA, in fact, only half as effective (d = 0.61 and 0.65).

On reading outcomes, training programs that were long-lasting yielded a significantly smaller effect size than shorter training programs as shown in Table 3. Effect sizes for the three shorter time blocks did not differ. The same pattern was evident on spelling outcomes.

These findings run counter to the expectation that more extensive training in PA should enable children to acquire superior phonemic awareness with stronger benefits for reading and spelling. These findings suggest that PA training does not need to be lengthy to exert its strongest effect on reading and spelling. However, caution is needed in drawing conclusions. There are various reasons why effect sizes might have been smaller when training was extensive. Perhaps the goals of instruction were more complex and harder to achieve. Or perhaps the students who received extended training were harder to teach. Alternatively, perhaps shorter instruction is better. The value of PA instruction may be to initiate insight into the alphabetic system. Adding further nuances or complexities may erode learning by producing confusion or boredom. In sum, the optimum length of PA training remains an issue needing further research.

Classroom teachers are the primary purveyors of reading instruction so, it is important to verify that they can teach PA effectively. Results of the analysis of phonemic awareness outcomes (see Table 2) showed that the effect size produced by classroom teachers was large (d = 0.78) although not as large statistically as that produced by others, consisting mainly of researchers (d = 0.94). This is not surprising, given that researchers were the ones who devised the training procedures in all of the studies.

PA training delivered by teachers transferred to reading and spelling. In the case of reading outcomes, the effect size associated with classroom teachers was significantly smaller (d = 0.41) than the effect size of researchers (d = 0.64). Of course, in these studies, neither teachers nor researchers intervened and helped children apply their PA skills in the reading transfer tasks. If transfer occurred, it was unassisted. This contrasts with normal classroom operations where teachers not only teach phonemic awareness but also teach children how to apply it in their reading and provide practice doing this. Under these circumstances, much more transfer to reading would be expected.

In the case of spelling outcomes, Table 4 reveals that effect sizes associated with classroom teachers were significantly greater than effect sizes associated with researchers (d = 0.74 vs. 0.51). However, the researcher effect size may have been depressed by the disproportionate presence of disabled readers in this category. When disabled readers were removed from the database, the effect sizes did not differ (see below).

There were only seven studies that used computers to teach PA. Ten treatment-control comparisons were derived from these studies. From PA outcomes in Table 2, it is apparent that computers produced a moderately strong effect size on the acquisition of PA (d = 0.66) although it was significantly less than the effect size for other forms of instruction (d = 0.89). The phonemic awareness that children learned from computers transferred and improved their reading performance on the immediate posttest (d = 0.33), but computers did not improve reading as much as other forms of PA instruction (d = 0.55). In contrast to the effects on reading, computer instruction exerted no significant effect on spelling outcomes (d = 0.09). One reason is that most of the computer comparisons involved disabled readers whose spelling performance did not benefit from PA training. From these findings the Panel concludes that computers are effective for teaching PA and for promoting transfer to reading, but they may be ineffective for teaching spelling to disabled readers.

Characteristics of Students

Some of the studies in the database targeted younger students at risk for future reading problems and older students classified as disabled readers. Both groups have been found to exhibit excessive difficulty manipulating phonemes in words (Bradley & Bryant, 1983; Juel, Griffith, & Gough, 1986; Juel, 1988). PA training programs were designed to remediate these readers' PA problems. Three types of readers were coded in the database: at-risk, disabled, and normally progressing readers. A comparison of phonemic awareness outcomes across the three groups revealed that although effect sizes were moderate to large in all cases, they were signficantly smaller for disabled readers (d = 0.62) than for at-risk (d = 0.95) and normally progressing readers (d = 0.93). This suggests that it was harder to improve PA in reading disabled students than in nondisabled students, perhaps because the disabled readers were older and relatively more advanced in PA skills with less room for gains than the younger beginning-level readers. Also it was the case that disabled readers were taught more advanced forms of PA (i.e., segmenting and blending with letters) than the younger students. At-risk readers were found to gain as much from PA training as normally developing readers. This indicates that having low PA when training began did not hinder at-risk readers in acquiring PA.

One might expect this pattern to be replicated on reading outcomes. However, Table 3 reveals that at-risk children showed bigger transfer effects in their reading (d = 0.86) than normal and disabled students whose effect sizes were equivalent (d = 0.47 for normals and d = 0.45 for disabled). Effect sizes on followup reading tests showed the same pattern except that the effect size for at-risk students was even larger (d = 1.33), while the effect sizes of the other two groups were smaller (d = 0.30 for normals and 0.28 for disabled). These findings indicate that PA training gives at-risk students a bigger boost in reading than it gives normals or disabled readers.

The effect of PA training on spelling outcomes differed among the three reader groups. Effect sizes were large and similar for at-risk (d = 0.76) and normal readers (d = 0.88). However, as indicated above, the effect size was much smaller, in fact, not significantly different from zero for disabled readers (d = 0.15). These findings show that PA training is not effective for improving disabled readers' spelling skills, perhaps because their spelling skills are much harder to remediate than their reading skills. In contrast, PA training was found to transfer to spelling in at-risk and normally progressing readers, indicating that PA training does benefit spelling in nondisabled readers.

The Panel also examined the effects of PA training at various grade levels: preschool, kindergarten, 1st grade, and 2nd through 6th grades. From PA outcomes in Table 2, it is evident that preschoolers showed a very large effect size in acquiring PA (d = 2.37). However, only two cases contributed to this value, making it less reliable. The effect on PA outcomes in kindergarten (d = 0.95) was significantly larger than the effect in 1st grade (d = 0.48) and in 2nd through 6th grades (d = 0.70). The latter two effect sizes did not differ. These findings indicate that younger students gained the most PA, not surprisingly since they started out with the least PA.

Effect sizes for reading outcomes in Table 3 reveal that PA training transferred to reading to a similar extent for kindergartners, 1st graders, and 2nd through 6th graders (ds from 0.48 to 0.49). The effect size for preschoolers was much larger (d = 1.25). The same pattern was not apparent on spelling outcomes, as evident in Table 4. Transfer of PA training to spelling was greater among kindergartners (d = 0.97) than among 1st graders (d = 0.52). There was no transfer to spelling among the 2nd through 6th graders for whom the effect size did not differ from zero (d = 0.14). (Spelling was not measured in the preschool studies.) The absence of an effect on spelling among the older children arose primarily because the majority of the cases in 2nd through 6th grades (78%) consisted of disabled readers who failed to show transfer effects from PA training to spelling (see below).

The Panel examined the relationship between the socioeconomic status of students across studies and the size of effects produced by PA training. As evident for PA outcomes in Table 2, low and mid-to-high SES levels did not differ, and both levels showed large effect sizes in acquiring PA. However, transfer to reading and spelling was significantly greater among among mid-to-high SES than among low SES students (see Tables 3 and 4). It might be noted that most studies of disabled readers did not report the students' SES; so, disabled reader effect siizes did not contribute to SES effect size calculations.

The NRP database included many studies conducted in English-speaking countries as well as a smaller number of studies conducted in countries speaking languages other than English. A comparison of effect sizes revealed that PA training exerted a larger impact on the acquisition of PA by English-speaking students (d = 0.99) than by the non-English students (d = 0.65). Transfer to reading outcomes was also greater for English students (d = 0.63) than for others (d = 0.36) on the immediate test but not the followup test. However, there were no differences in effects sizes on spelling outcomes.

A possible reason for the absence of effects on spelling is that most of the studies involving disabled readers were in the pool of English studies. This may have suppressed the English effect size in spelling. To check on this, effect sizes were recalculated with the reading-disabled (RD) comparisons removed (see below). Results confirmed suspicion; they changed from no effect on spelling to a significant effect favoring English (d = 0.95) over non-English (d = 0.51).

One intriguing reason for the larger effect sizes in English may be that the English writing system is not as transparent in representing phonemes as it is in the other languages; so, explicit training may make a bigger contribution to clarifying phoneme units and how they link to graphemes in words for English-speaking students.

Analysis of Moderator Effects With Disabled Readers Removed From the Database

In the analysis of effects associated with the three types of readers, effect sizes were significantly smaller for disabled readers than for at-risk and normal readers on two outcomes, phonemic awareness and spelling. In fact, on the spelling outcome, no significant effect of PA training was detected for disabled readers. Moreover, the pool of spelling effect sizes for disabled readers was homogeneous, indicating that no further analysis of moderator variables was needed to locate cause and allowing us to conclude that PA training does not improve spelling in disabled readers.

44

In the NRP database, there were 17 comparisons involving disabled readers (18% of the total comparisons). The Panel worried that conclusions about how moderator variables regulate the impact of PA training on phonemic awareness and spelling outcomes might be different if cases involving disabled readers were removed from the database. As discussed above, in our analysis of English and non-English studies, findings changed for spelling outcomes with reading disabled cases eliminated. This was because the distribution of disabled reader cases was uneven, with most cases falling in the English pool of effect sizes. There were other moderator variables with an uneven distribution of disabled readers across levels as well. Disabled readers were older (mostly in grades 2 through 6), they tended to receive PA instruction involving multiple skills taught with letters, the instruction was individualized, it tended to be lengthy (over 19 hours), and researchers or computers rather than teachers were most often the trainers.

To examine whether findings involving these moderators would be different without disabled readers, effect sizes were re-analyzed after removing disabled reader comparisons from the database. The following specific moderator variables were re-analyzed: PA skills taught, use of letters, grade, language, training unit, teachers vs. others as trainers, and length of training. Computer effects were not re-analyzed because there were too few cases.

Findings involving spelling outcomes were altered for several moderators when disabled readers were removed. Findings involving PA outcomes were altered for one moderator. However, findings were not altered at all in the analyses of reading outcomes. Results are given in Table 5 (Appendix D).

Comparison of the number of cases contributing effect sizes to spelling outcomes with and without disabled readers (Tables 4 vs. Table 5) reveals that the numbers dropped substantially in the following categories: three or more PA skills taught (drop from ten to three cases), letters manipulated (from 27 to 17 cases), individual instruction (from 14 to 8 cases), small group instruction (from 20 to 15 cases), training lasting 20 to 75 hours (from 18 to 9 cases), researcher as trainer (from 30 to 20 cases), 2nd through 6th graders (from 8 to 0 cases), English language (from 32 to 22 cases). The same comparison for PA outcomes (Table 2 vs. Table 5) reveals that in the category of letters manipulated, the number dropped from 39 to 25 cases. Declines in the other categories listed in Table 5 were minimal. This verifies that disabled readers were unevenly distributed across levels of these moderators. The SES variable was not affected and hence not re-analyzed because most studies involving disabled readers did not report the SES level of the readers.

In all but one analysis of spelling outcomes, the pattern of effect sizes changed when disabled readers were removed from the database. PA teaching that focused on one or two skills was no longer superior to multiple PA skill teaching. (However, note in Table 5 that there were only three cases left in the multiple skills category, raising doubt about the reliability of this effect size.) Small group instruction no longer produced better transfer to spelling than individual instruction. Training periods lasting 20 or more hours were no longer less effective than shorter training periods. Classroom teachers no longer differed from researchers in facilitating transfer to spelling. In the analysis of spelling outcomes across grades, the 2nd through 6th grade category had no comparisons to contribute effect sizes. The loss of cases in the upper grades shows that disabled readers clearly dominated effect sizes in this category. The greater effect of PA training on spelling among kindergartners than 1st graders remained the same.

There were two moderators that did not differentially influence spelling or PA outcomes when the whole database was analyzed; but when disabled reader effects were removed, significant differences appeared. As evident in Table 5, language now impacted spelling effect sizes, with English-speaking students benefiting more from PA training than non-English-speaking students. Also, letter use now impacted phonemic awareness effect sizes such that children who manipulated letters acquired more PA than children who did not. Removal of disabled readers rendered findings for these moderators consistent across all three outcomes. That is, language exerted the same impact on PA, reading, and spelling outcomes, with English producing larger effects than non-English. Also letter use exerted the same impact on PA, reading and spelling, with letter manipulation producing larger effects than no letters.

In sum, these findings support the following conclusions. PA training does not improve spelling in disabled readers, but it does improve spelling in normally developing readers below 2nd grade and children at risk for future reading problems. Among nondisabled readers, the benefit to spelling is positive and does not depend on whether one or two or multiple PA skills are taught, whether instruction is delivered to individuals or to small groups, how long training lasts, or whether teachers or researchers are the trainers. However, the benefit to spelling among nondisabled readers does depend upon the language, with PA training in English exerting a bigger impact on spelling than PA training in other languages.

Regarding the acquisition of phonemic awareness by nondisabled readers, our findings support the conclusion that PA training is more effective when it is taught by having children manipulate letters than when manipulation is limited to speech.

It is important to note that the pattern of effect sizes on reading outcomes remained unchanged when comparisons involving reading disabled students were removed. Specifically, teaching one or two PA skills still resulted in larger effect sizes on reading than teaching a multitude of PA skills. Small groups still produced superior transfer to reading than individual instruction. Lengthy training periods still yielded smaller effects on reading than shorter training periods. These findings serve to sustain our conclusions about the influence of moderators on reading outcomes.

Design Features

Studies in the database varied in methodological rigor. The Panel examined some of these properties to see whether design weaknesses inflated effect sizes.

Studies varied in whether or not subjects were randomly assigned to treatment and control groups. In some cases, nonrandom, nonequivalent groups were assigned to treatment and control conditions. In some cases, group assignment involved matching individual children on the basis of similar test scores. Effect sizes for the three assignment types were determined (see Tables 2, 3, and 4 in Appendix C). Comparison of PA outcomes revealed very similar effect sizes that did not differ statistically and ranged from 0.83 to 0.92. Comparison of reading outcomes revealed that the effect size for randomly assigned groups ($\underline{d}$ = 0.63) was significantly greater than the effect size for nonequivalent groups ($\underline{d}$ = 0.40). However, the opposite was found on spelling outcomes, with nonequivalent groups showing a significantly larger effect size ($\underline{d}$ = 0.86) than random groups ($\underline{d}$ = 0.37). These findings show that larger effect sizes in our database did not consistently arise from weaker designs involving nonequivalent groups. Moreover, average effect sizes for the most rigorous assignment procedure, random assignment, ranged from low-moderate to large.

Some researchers in the database administered fidelity checks to ensure that trainers adhered to prescribed training procedures, whereas other researchers did not, or at least did not report, doing this. A comparison revealed that significantly larger effect sizes arose in studies not checking for fidelity than in studies checking for fidelity. This was true across all three outcome measures (see Tables 2, 3, and 4 in Appendix C). Although weaker studies involving lack of fidelity checking were associated with larger effects, fidelity studies nevertheless yielded significant effects that were moderate in size. This verifies that lack of rigor in fidelity checking does not explain effect sizes in the NRP database.

Bus and van Ijzendoorn (1999) reported an unexpected finding in their PA meta-analysis, that studies using treated control groups yielded larger effect sizes than studies using untreated control groups. This finding was examined in the present meta-analysis. Results were mixed. On PA outcomes, the two types of control groups did not yield significantly different effect sizes. On reading outcomes, they did, with studies using treated controls showing larger effects than those using untreated controls, consistent with Bus and van Ijzendoorn's finding. On spelling outcomes, studies with untreated controls showed larger effects than studies with treated controls, the reverse pattern.

The foregoing results emerged from an analysis of all the studies. However, these studies varied in many respects besides the type of control group they used. In the NRP database, there were eight studies that compared PA training to both a treated control group and an untreated control group. In limiting the analysis to these studies, the Panel found that, out of 20 comparisons, ten showed bigger effects in cases using treated controls and ten showed bigger effects in cases using untreated controls across the three outcome measures. Thus, the picture arising from this analysis was mixed.

Although the findings reveal no clear pattern favoring treated or untreated control groups, the fact that studies using untreated controls did *not* uniformly yield larger effect sizes serves to challenge the commonly held belief that untreated control groups always yield larger effects. It is not the case that Hawthorne effects always prevail. Other factors appear to influence outcomes as well. Perhaps Hawthorne effects are more characteristic of older participants with better developed metacognitive sensitivities.

Among studies in the NRP database, samples included as few as nine students or as many as 383 students. To examine whether effects differed as a function of sample size, the studies were divided into blocks of approximately equal numbers of cases. Outcomes reported in Tables 2 to 4 reveal that larger effect sizes tended to occur in the smaller samples, whereas the smallest effect sizes occurred in the largest samples. This is consistent with meta-analytic findings in general (Johnson & Eagley, in press). The fact that effect sizes were significantly greater than zero even in the largest samples shows that the PA training effects observed did not arise primarily from the weaker studies with small samples.

Recently Troia (1999) published a critique of phonemic awareness training studies. He identified several criteria to assess methodological rigor and applied these criteria to 39 PA training studies of which 29 were in the NRP database. (The remaining studies did not assess reading as an outcome so were not among the studies considered.) The Panel incorporated his summary ratings into the NRP database and examined the relationship between these evaluations and effect sizes. Troia devised two measures and applied them to evaluate the internal validity separately from the external validity of studies: the percentage of criteria met and the number of critical flaws. Also he ranked the studies to indicate their overall methodological rigor. The Panel's purpose was to consider and rule out the possibility that effects of PA training were limited primarily to studies that were the least rigorous. Comparisons were grouped into blocks of three or four in order to reveal effect sizes at the various levels of rigor.

The findings are reported in Appendix E—Table 6 for PA outcomes and Table 7 for reading outcomes. Both tables reveal that effect sizes were significantly greater than zero across all blocks on all five measures. This shows that significant effect sizes were not limited to the weakest studies.

In Table 6, reporting effects of PA training on PA outcomes, it is apparent that across all five measures the largest effect sizes occurred for the blocks reflecting the most rigor. This shows that the best designed studies produced the largest effect sizes on the acquisition of PA.

In Table 7, reporting effect sizes for reading outcomes, the same pattern is evident but is not quite as strong. The effect size associated with the most rigorous level is close to the strongest, if not the strongest, effect size on four of the five measures: the two internal validity measures, the external validity critical flaws measure, and the overall rigor ranking. On the remaining measure, percent of external validity criteria met, the effect size is moderately strong though less so than the largest effect size. This evidence indicates that the better designed studies tended to produce stronger transfer effects in reading than the weaker studies.

In sum, although Troia (1999) finds fault with PA training studies, his findings do not undermine claims about the effectiveness of PA training for helping children learn to read. Troia's concluding plea, that researchers maintain high standards in designing their studies, is supported by Panel findings that show that researchers stand a better chance of obtaining sizeable effects when they design strong studies than when they design weak studies threatened by violations to internal and external validity.

One final characteristic of studies examined was the year of publication. From Tables 2 and 3, it is apparent that there was one period in which a spate of PA training studies was published, from 1991 to 1994. Over twice as many studies were published during this period as during the other periods. The 1991 to 1994 studies also tended to yield larger effect sizes on PA and reading outcomes than studies in time periods before or after this. Why this occurred is not clear.

Discussion

Summary of Findings

To summarize results of the meta-analyses, the Panel examined 96 cases, each comparing a treatment group that received PA training, to a control group that received an alternative form of instruction or no special instruction; they examined effects on three main outcome variables, PA, reading, and spelling.

PA training was found to be very effective in teaching phonemic awareness to students. Effect sizes were large immediately after training (d = 0.86), and they remained strong over the long term (d = 0.73). PA training succeeded in teaching children various ways to manipulate phonemes, including segmentation, blending, and deletion. PA training was effective in teaching PA skills across all levels of the moderator variables examined.

PA training improved children's ability to read and spell in both the short and the long term. The effect size was moderate following training on reading (d = 0.53) and on spelling (d = 0.59). Tests of word reading, pseudoword reading, and reading comprehension all yielded statistically significant effect sizes on both experimenter-devised tests as well as standardized tests. Few instances occurred in which moderator variables reduced effect sizes to chance levels, and these were limited to spelling outcomes. Whereas PA training exerted strong effects on reading and spelling, it did not impact children's performance on math tests. This indicates that halo/Hawthorne effects did not explain findings and that training effects were limited to the targeted domain.

Several moderator variables were found to influence children's acquisition of phonemic awareness. PA training programs varied in whether children were taught to manipulate phonemes in one, two, or multiple ways, and in the type of phoneme manipulations taught, segmenting, blending, deleting, identifying, or categorizing phonemes, or manipulating onsets and rimes. Properties of the training procedures exerted an impact. Programs that focused on teaching one or two PA skills yielded larger effects on PA learning than programs teaching three or more of these manipulations. Instruction that taught phoneme manipulation with letters helped children acquire PA skills better than instruction without letters. Facilitation from letters was observed among at-risk readers and normally developing readers below 2nd grade. It was not possible to assess the contribution of letters among disabled readers because most studies used letters to teach PA to disabled readers.

Teaching children in small groups produced larger effect sizes on PA acquisition than teaching children individually or in classroom-size groups. Classroom teachers produced large effect sizes, indicating that they were very successful in teaching PA to students, although researchers produced somewhat larger effects. Computers also taught PA effectively. The length of training influenced PA acquisition. Effect sizes were larger when PA instruction lasted from 5 to 18 hours than when either less or more time than this was spent.

Characteristics of students influenced how much phonemic awareness they acquired from training. Disabled readers showed smaller effect sizes than at-risk students or normally progressing readers, indicating that PA was harder for disabled readers to learn. Also students in the lower grades, namely preschool and kindergarten, showed larger effect sizes in acquiring PA than children in 1st grade and above. SES exerted no differential impact on learning PA. However, the language spoken by the children did. English-speaking children showed larger effects of training on PA acquisition than children learning in other languages.

These moderator variables also influenced how much transfer to reading and spelling resulted from PA training. The type of test used to measure reading and spelling influenced effect sizes that were larger on experimenter-devised tests than on standardized tests measuring real word reading and spelling. Effect sizes did not differ on experimenter-devised and standardized pseudoword reading tests.

Properties of training procedures influenced the extent of transfer to reading. Teaching that focused on one or two types of PA manipulations yielded larger effect sizes than teaching three or more PA skills. Teaching children to manipulate phonemes using letters produced bigger effects than teaching without letters. Blending and segmenting instruction showed a much larger effect size on reading than multiple-skill instruction did. Small group instruction produced larger effect sizes on reading than individualized instruction or classroom instruction. Length of training exerted an influence as well, with the lengthiest training associated with the smallest effect size. Classroom teachers provided PA training that was effective in promoting transfer to reading although the effect size of teachers was smaller than the effect size of other trainers. PA training on computers transferred to reading as well.

Characteristics of learners influenced the extent that PA training transferred to reading. Effect sizes on reading were large for at-risk readers while they were moderate for disabled and normally developing readers. Preschoolers exhibited a much larger effect size on reading than did the other grade levels whose effect sizes did not differ. SES made a difference, with mid-to-high SES associated with larger effects than low SES. Also larger effect sizes were evident in reading for English-speaking children than for children speaking other languages.

Analysis of moderator variables as they affected spelling outcomes was complicated by the fact that PA training did not help disabled readers improve in spelling and the pool of spelling effect sizes for disabled readers was homogeneous, indicating that further analyses using moderators was not necessary to explain the result. The effects of moderators were re-analyzed with disabled readers removed from the database. Conclusions regarding the effects of moderator variables on spelling outcomes thus centered on the nondisabled readers.

The only characteristic of PA training that influenced spelling outcomes for nondisabled readers was the use of letters. Children who were taught to manipulate phonemes with letters benefited more in their spelling than children whose manipulations were limited to speech. Whether instruction focused on one or two skills or on multiple skills did not influence spelling in nondisabled readers. Instruction delivered to individuals was as effective as instruction delivered to small groups, and both were more effective than classroom-size groups. The length of training exerted no differential impact on spelling outcomes. Whether the trainer was a teacher or a researcher made no difference. Characteristics of learners did make a difference. Kindergartners benefited more in their spelling than did 1st graders. Students classed as mid-to-high SES showed a larger effect size in spelling than low SES students. PA training in English produced a larger effect on spelling than PA training in other languages.

Features of the design of experiments were related to effect sizes. Findings indicated that rigorous designs yielded strong effects. The majority of the studies used random assignment, and their effect sizes on PA and reading outcomes ranged from moderate to large. About one-third of the studies checked on whether trainers remained faithful to treatment procedures. Effect sizes in these studies were significant and moderate in size. Some studies compared PA treatment groups to control groups that were given some other treatment while other studies used untreated control groups. Neither type of control group consistently produced larger effect sizes. Failure to find larger effects for untreated than for treated control groups indicates that Hawthorne effects did not inflate effect sizes. Studies using smaller samples of children tended to have larger effect sizes than studies using larger samples, a finding consistent with other meta-analyses. However, even in the largest samples, effect sizes were positive and significant.

The Panel also assessed the relationship between methodological rigor and effect size by applying Troia's (1999) criteria to the NRP studies. On PA outcomes, the best designed studies produced the largest effect sizes on all five measures of rigor. On reading outcomes, effect sizes associated with the most rigorous level were close to the largest, if not the largest, effect sizes on four out of five measures: two internal validity measures, one external validity measure, and the overall ranking of rigor. This indicates that the better designed studies produced larger transfer effects in reading than the weaker studies. In sum, findings show that larger effect sizes did not arise mainly from weaker studies that were flawed by threats to internal and external validity.

Interpretations and Issues

Results of the experimental studies allow the Panel to infer that PA training was the cause of improvement in students' phonemic awareness, reading, and spelling performance following training. These findings were replicated multiple times across experiments and thus provide solid support for causal claims. However, results of the analysis of moderator variables rest on more tentative ground. Assessing features of the studies that were associated with stronger or weaker effect sizes is at root a correlational endeavor and thus precludes strong inferences about cause. The primary difficulty is that a third unknown factor may lie in the background explaining the relationships observed. Although findings are suggestive, any conclusions must remain tentative because multiple explanations are possible. In this section, potential misinterpretations of the findings and issues needing further attention from researchers are considered.

The studies in the NRP database included investigations of children at risk for future reading problems as well as children low in SES. However, contrary to the common view that the criteria for identifying at-risk readers includes being economically disadvantaged, authors of the studies investigating at-risk readers did not uniformly require them to be low in SES. In fact, of the cases investigating at-risk readers, only 27% were low in SES while 37% were middle-to-high SES, and the SES of the remainder was not specified. At risk was defined by low phonemic awareness in 77% of the cases. In defense of these studies, research findings show that one of the two best predictors of reading success is phonemic awareness (Share et al., 1984), so selecting at-risk readers by measuring their PA makes sense. However, because the training targeted this skill, large effect sizes may be less surprising.

The fact that studies in the NRP database departed from the common conception of what it means to be at risk serves to reconcile discrepancies between results for at-risk readers and results for low SES readers. The Panel found that at-risk children showed large effect sizes in acquiring PA (d = 0.95) and in transferring these skills to reading (d = 0.86) and spelling (d = 0.76). Low SES children also showed large effect sizes in phonemic awareness (d = 1.07) and spelling (d = 0.76), but only a moderate effect size in reading (d = 0.45). Smaller effect sizes in reading among low SES children than among at-risk children is explained by the fact that the majority of the at-risk children were not low in SES. Based on these findings, one would expect at-risk children who are both low PA and low SES to exhibit large gains in PA and spelling as a result of PA training but to exhibit moderate gains in reading.

It is noteworthy that low SES children were found to benefit as much from PA training as middle-to-high SES children in acquiring phonemic awareness. This runs counter to Dressman (1999) who argues that low SES children will exhibit low PA in research studies because their phonological systems differ from that of testers and because they suffer from inhibition when tested by sociolinguistically foreign researchers. Dressman bases his expectations on studies showing that low SES children perform more poorly on PA tests than middle-class children. He ignores evidence examining how much low SES children gain in PA when they receive training. According to the NRP findings, low SES children can benefit as much from training as middle-to-high SES children, despite being phonologically or culturally different from the trainers.

One very striking finding was that in contrast to at-risk and normally developing readers, disabled readers' spelling did not benefit at all from PA training. Various reasons for this can be entertained. Other studies have found that disabled readers have special difficulty learning to spell (Bruck, 1993). Perhaps processing difficulties associated with being reading disabled make spelling especially hard to learn. Alternatively, perhaps PA training fails to help older disabled readers with their spelling because the types of words that are spelled in higher grades require knowledge of spelling patterns rather than phonemic segmentation and knowledge of individual letter-sound correspondences. Effects of PA training on spelling may be limited to less complex words that are more phonemically transparent, those taught to beginning readers.

According to NRP findings, children who received training that focused on one or two PA skills exhibited stronger PA and stronger transfer to reading than children who were taught three or more PA skills. Various explanations might account for the difference. Perhaps focused instruction resulted in more students mastering the skills that were taught. Perhaps teaching multiple skills created some confusion about which manipulations to apply in the reading transfer tasks, or perhaps it obscured children's grasp of the alphabetic principle. Clarifying why multiple skills instruction might limit children's gains in PA and reading needs further study. However, the findings suggest that when multiple

PA skills are the objective, it is prudent to teach one at a time until each is mastered before moving on to the next, and to teach students how each skill applies in reading or spelling tasks.

More important than the number of PA skills to teach is the question of which skills should be taught to children. In all of the studies, children were given PA instruction that was considered appropriate for their level of literacy development. The manipulations taught to preschoolers were quite different from the manipulations taught to older students. Easier PA tasks were taught to younger children or to less mature readers while harder PA tasks were taught to older readers. Factors making PA tasks easy or difficult include the type of manipulation applied to phonemes, the number and phonological properties of phonemes in the words manipulated, whether the words are real or nonwords, and whether letters are included. To illustrate, the following tasks are ordered from easy (1) to difficult (6) based on findings of Schatschneider, Francis, Foorman, Fletcher, and Mehta (1999):

1. First-sound comparison—identifying the names of pictures beginning with the same sound
2. Blending onset-rime units into real words
3. Blending phonemes into real words
4. Deleting a phoneme and saying the word that remains
5. Segmenting words into phonemes
6. Blending phonemes into nonwords.

In the illustrative studies described below, tasks that are appropriate to teach at different grade and reader levels can be seen. The final decision about which PA manipulations to teach should take account of several factors, not only task difficulty, but also whether or not students can already perform the manipulations being taught as determined by pretests, and the use that students are expected to make of the PA skill being taught. The reason to teach first-sound comparisons is to draw preschoolers' or kindergartners' attention to the fact that words have sounds as well as meanings. A reason to teach phoneme segmentation is to help kindergartners or 1st graders generate more complete spellings of words. The reason to teach phoneme blending is to help 1st graders decode words.

One surprising finding in the analysis involved the relationship between training time and outcomes. Effect sizes were larger when PA instruction lasted between 5 and 18 hours than when either less or more time was spent training students. However, caution is needed in interpreting this finding because multiple explanations are possible. Perhaps the goals of instruction were more complex in longer programs. Perhaps the students receiving instruction were harder to teach. Perhaps spending many hours in PA training deprived students of the reading instruction benefiting control groups. Perhaps PA instruction is valuable mainly in helping children achieve basic alphabetic insight. Going beyond this by adding further nuances or complexities may erode learning by producing confusion or boredom. These are only some of the possible reasons why longer training sessions might have produced smaller effect sizes. Questions regarding the optimum length of PA training and factors determining optimum length invite further research. However, two conclusions seem self-evident: that length of training should be regulated by how long it takes students to acquire the PA skills that are taught and that the NRP findings should not be translated into any prescriptions regarding how long teachers should spend teaching PA.

One important moderator variable that was not considered in the analysis is dialect because none of the studies paid attention to this variable. However, regional differences at the phonemic level of language are likely to be important. For example, vowel phoneme categories are not the same across the United States. Some dialects make more phonemic distinctions among vowels than other dialects. Vowels in the three words, marry, Mary, and merry are pronounced identically in some areas of the West but differently in some areas of the East. As a result, no generalizations about these vowel phonemes will suit everyone receiving PA instruction. Another dialectal difference involves preserving or deleting the final consonants in words, for example, past-tense markers such as the /t/ in looked. More research on the impact of dialectal variations on PA learning is needed. The fact that regional phonemic variations exist means that teachers implementing PA training programs need to be aware of their students' dialects and whether they deviate from the phonological systems that are assumed in the programs. Ignoring deviations is likely to undermine the credibility of the instruction.

Another variable related to students' phonological systems but neglected in the analysis is whether English is the first or second language of students. The problem here is that phonemes in English may not be phonemes in ESL students' first language. To understand this requires distinguishing between phonemes and phones. Phonemes are the smallest units in speech that signal a difference in meaning to a listener who knows the language. Phones are also the smallest units in speech but are described by acoustic and articulatory properties. To perceive phonemes, speakers use categories that were constructed in their minds when they learned their particular language. In contrast, phones are defined by their physical properties. Phonemes are broader categories that may include several phones, called allophones, differing in their articulatory features. Even though the allophones differ, speaker/listeners process them as the same phoneme. For example, the initial sounds in chop and shop are articulated differently, so they are two different phones. To an English speaker, they are also different phonemes, because substituting one for the other signals a different word. However, to a speaker of Spanish, the two different phones are the same phoneme. The change in articulation does not signal a different word in Spanish. The speaker either fails to notice the difference or perceives it as a slightly different way of pronouncing the same word. Another example is that Chinese and Japanese speakers process /l/ and /r/ as the same phoneme in English words.

The distinction between phonemes and phones may seem trivial, but it is not. If teachers have students who are learning English as a second language, they need to realize that their students are almost bound to misperceive some English phonemes because their linguistic minds are programmed to categorize phonemes in their first language, and this system may conflict with the phoneme categorization system in English. Their confusions will be most apparent when they select letters to spell unfamiliar words. If they know Spanish, they may select CH when they should use SH. If they know Japanese or Chinese, they may confuse L and R. When teachers teach PA, they need to be sensitive to these sources of difficulty faced by their ESL students.

The Role of PA in Reading Acquisition Processes

Findings of the meta-analyses show that PA training benefits the processes involved in reading real words, pseudowords, and text reading. It also benefits spelling skills in normally progressing readers below 2nd grade and in beginners at risk for developing reading problems. There are several reasons why PA training is thought to help children learn to read and spell.

The English writing system is alphabetic. Breaking the code entails figuring out how graphemes represent phonemes. These relationships, though systematic, are variable across word spellings. The same letters may symbolize more than one phoneme, and single phonemes may be represented by alternative graphemes. The vowels are especially variable. This lack of transparency makes it harder for beginners to figure out the system without help.

Speech is seamless and has no breaks signaling where one phoneme ends and the next begins. Also, phonemes overlap and are coarticulated, which further obscures their separate identities. Another barrier to developing PA is that speakers focus their attention on the meanings of utterances, not on sounds. Unless they are trying to learn an alphabetic code, there is no reason to notice and ponder the phonemic level of language. These facts explain why beginners have difficulty acquiring PA and why they benefit from explicit instruction in PA.

An essential part of the reading process involves learning to read words in various ways (Ehri, 1991, 1994). Because phonemes in words correspond to graphemes in the English writing system, all of these ways of reading words are easier to acquire when beginners possess PA. Phoneme identity is needed to attach phonemes to letters for reading and spelling words. The skill of blending is needed to decode unfamiliar words. Being able to segment and blend onsets and rimes in words helps children read unfamiliar words by analogy to known words. Phonemic segmentation helps children remember how to read and spell words because it helps them distinguish the phonemes that are bonded to graphemes when a word's written form is retained in memory. When unfamiliar words are read in text, students may apply decoding skills, or they may combine grapheme-phoneme cues with meaning cues to derive the word (Tunmer & Chapman, 1998).

It is important to note that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping children understand and use the alphabetic system to read and write. This is why including letters in the process of teaching children to manipulate phonemes is important. PA training with letters helps learners determine how phonemes match up to graphemes within words and thus facilitates transfer to reading and spelling.

It is important to recognize that children will acquire some phonemic awareness in the course of learning to read and spell even though they are not taught PA explicitly. The process of learning letter-sound relations and how to use them to read and spell enhances children's ability to manipulate phonemes. This is indicated by evidence that people who do not learn to read in an alphabetic system do not develop PA (Mann, 1987; Morais et al., 1987; Read et al., 1987). It is also indicated by the fact that, in many of the studies reviewed, control groups showed improvement in phonemic awareness from pretests to posttests, very likely because of the reading and writing instruction they received in their regular classrooms. However, the extent of PA needed to contribute maximally to children's reading development does not arise from incidental learning or instruction that is not focused on this objective. This is indicated by the finding that children receiving explicit training in PA gained much more PA and reading skill than children in the control groups.

It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness; therefore, PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling while others may know only a few letters and have no reading skill. The nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA prior to beginning PA instruction. This will indicate which children need the instruction and which do not; which children need to be taught rudimentary levels of PA, for example, segmenting initial sounds in words; and which need more advanced levels involving segmenting or blending with letters.

In the rush to teach phonemic awareness, it is important not to overlook the need to teach letters as well. The NRP analysis showed that PA instruction was more effective when it was taught with letters. Using letters to manipulate phonemes helps children make the transfer to reading and writing. However, teaching children all the letters of the alphabet is not easy, particularly when they come to school knowing few of them. There are 52 capital and lower-case letter shapes, names, and sounds to learn. The shapes of many letters are similar, and, therefore, easily confused with one another. Letter learning requires retaining shapes, names, and sounds in memory and, in fact, overlearning them so that letters can be processed automatically in reading and writing words (Adams, 1990). Thus, to ensure that instruction in phonemic awareness is effective, it needs to include instruction in graphemes as well as instruction in the connections between graphemes and phonemes to read and spell words.

In addition to teaching PA skills with letters, it is important for teachers to help children make the connection between the PA skills taught and their application to reading and writing tasks. In most of the studies reviewed, researchers did not do this when they presented the transfer tasks to students following training. Despite this, significant and sizable transfer effects were observed. In a study by Cunningham (1990), who did examine application effects, students in one group not only were taught to segment and blend but also were shown how to apply these skills in reading words. Another group received the same PA training but not the application training. Effect sizes on reading outcomes were much larger when 1st graders received the application instruction than when they did not. This suggests that results of the NRP meta-analysis actually underestimate the magnitude of effects that would result if children received explicit instruction and practice in applying PA skills in their reading and writing.

It is important to note that when PA is taught with letters, it qualifies as phonics instruction. When PA training involves teaching students to pronounce the sounds associated with letters and to blend the sounds to form words, it qualifies as synthetic phonics. When PA training involves teaching students to segment words into phonemes and to select letters for those phonemes, it is the equivalent of teaching students to spell words phonemically, which is another form of phonics instruction. These methods of teaching phonics existed long before they became identified as forms of phonemic awareness training (Balmuth, 1982; Chall, 1967). Although teaching children to manipulate sounds in spoken words may be new, phonemic awareness training that involves segmenting and blending with letters is not. Only the label is new. Explicit instruction in the alphabetic principle necessarily includes attention to phonemes because these are the phonological units that match up to letters. According to NRP findings, it is likely that the inclusion of phonemic awareness training in phonics instruction is a key component contributing to its effectiveness in teaching children to read.

It is important to note that various approaches to beginning reading instruction may provide at least some phonemic awareness training although it may not be presented systematically or thoroughly enough to maximize its contribution to reading and writing. Whole language instruction that teaches students to invent spellings by detecting phonemes in words and representing them with letters offers a form of PA training. In Reading Recovery® (RR), students may acquire phonemic awareness through the spelling instruction they receive (Clay, 1985). Three studies in the database compared outcomes of standard whole language instruction, or RR instruction, to outcomes of the same instruction with PA training added (Castle et al., 1994; Hatcher et al., 1994; Iversen & Tunmer, 1993). Overall effect sizes were variable ranging from negative to large positive (see Appendix and illustrative studies below). One factor possibly limiting outcome differences between treatment and control groups is the extent to which control students acquired PA from the instruction they received. Although whole language programs and RR programs include some phonemic awareness training, findings of the NRP meta-analysis indicate that strengthening the training offered in spelling activities by making it more systematic, thorough, and explicit, is likely to improve these programs' success in helping children learn to read and spell.

Classroom Instruction in PA: Some Illustrations

NRP findings show that PA training programs implemented by teachers in classrooms are effective in teaching phonemic awareness to students, and this training boosts children's reading and spelling performance. To identify characteristics of programs that were used successfully by classroom teachers, the Panel examined a few illustrative studies selected from a total of 15 (Blachman, Ball, Black, & Tangel, 1994; Brady, Fowler, Stone, & Winbury, 1994; Brennan & Ireson, 1997; Bus, 1986; Haddock, 1976; Kennedy & Backman, 1993; Kozminsky & Kozminsky, 1995; Lie, 1991; Lundberg, Frost, & Petersen, 1988; McGuinness, McGuinness, & Donohue, 1995; O'Connor, Notari-Syverson, & Vadasy, 1996; Olofsson & Lundberg, 1983; Schneider, Kuspert, Roth, Vise, & Marx, 1997; Tangel & Blachman, 1992; Williams, 1980).

One 8-month-long, carefully structured program for kindergartners was developed and tested by Lundberg, Frost, and Peterson (1988). Twelve classroom teachers in Denmark taught children daily to attend to sounds in speech and to manipulate sounds through games and exercises that increased in difficulty as the year progressed. The program began with easy listening activities followed by rhyming exercises. Then kindergartners learned to segment sentences into words and to focus on the length of words in speech. Then words were analyzed into syllables. For example, children listened to a troll who spoke peculiarly, syllable by syllable, and they figured out what he said. Phoneme analysis was introduced in the 3rd month by having children identify phonemes in initial positions of words, mainly continuants and vowel sounds which are easy to stretch out and hold. The teacher helped children find the sounds by stretching them, for example, "Mmmmmark" or by repeating the stop consonants that cannot be held, for example, "T-T-T-Tom." Children also practiced adding and deleting phonemes from words. In the 5th month of the program, phoneme segmentation and blending were introduced, first with two-phoneme words and then longer words. Many of the activities were designed for children's enjoyment and consisted of dancing, singing, and other noncompetitive social games.

Teachers were trained in an inservice course that provided theoretical background as well as videotaped examples of training sessions. They practiced and refined the skills necessary to teach the program during the year prior to implementing it. Teachers of the control group followed the regular preschool program, which emphasized social and aesthetic aspects of development rather than cognitive and linguistic aspects. Treatment and control schools were located in geographically distant parts of Denmark.

The Danish program was adapted and tested by other researchers including Schneider et al. (1997) who taught PA to German kindergartners. His study included two experiments and a total of 22 teachers who taught PA in the treatment conditions. Control groups received the regular kindergarten curriculum. The second experiment was conducted to improve on the first. Teacher training was less extensive in the German study than in the Danish study. It lasted 2 months and included theoretical background and tutoring sessions in which teachers practiced the games and exercises and received feedback.

In both the Danish and German studies, training produced large effect sizes on the acquisition of phonemic awareness, ranging from 0.70 to 0.82. Effect sizes on reading outcomes were small to moderate when measured the following year in 1st grade: d = 0.19 (Denmark), d = 0.26 and 0.45 (Germany).

An adaptation of the Danish program was tested with English-speaking kindergartners by Brennan and Ireson (1997). However, only one teacher and her class of 12 students formed the PA treatment group, which was compared to one no-treatment control class. Although this is a weaker design yielding less reliable findings, the effect size was impressive. The impact of training on word reading was large, with an effect size of d = 1.17. This provides some evidence that the Danish program can be used effectively in American classrooms. A translation of the program has been published (Adams, Foorman, Lundberg, & Beeler, 1998). Whether teachers need further help beyond the manual to implement the program effectively with their students needs to be studied.

The Danish program did not include letter manipulation. However, the meta-analysis showed that when PA is taught with letters, it is more effective. A program for kindergartners that included letters was developed and tested by Blachman and her colleagues (Ball & Blachman, 1991; Blachman et al., 1994; Tangel & Blachman, 1992). Blachman et al. (1994) taught 10 teachers and their teaching assistants to deliver PA training to low-income, inner-city kindergartners. Children were taught in groups of four or five for 15 to 20 minutes per day, 4 times each week. The program lasted 11 weeks. The teachers were trained in seven 2-hour inservice workshops, during which they were taught a theoretical framework; they practiced instructional activities; and they asked questions about ways of implementing the program.

A key activity in Blachman et al.'s (1994) program was the "say it and move it" procedure. Children learned to move a blank tile down a page as they pronounced each phoneme in a word. After children practiced segmenting two- and three-phoneme words in this way, letter-sound correspondences were taught and they practiced segmenting the words with blank markers and letters. Additional segmentation activities were included such as moving markers into Elkonin boxes to represent phonemes in three-phoneme words. A variety of games was used to reinforce grapheme-phoneme correspondences. The control group in this study followed a traditional kindergarten curriculum that included instruction in letter names and sounds. Results of the study were very positive. Children receiving PA training outperformed controls on PA tasks, with an effect size of d = 1.83, and training transferred to reading, d = 0.65, and to spelling, d = 0.94.

Another program in the NRP set of studies was administered by teachers to small groups of older disabled readers. Williams (1980) developed and tested the ABDs program, which taught students ages 7 to 12 to segment and blend phonemes first in speech and then using letters. Children worked with a limited set of seven consonants and two vowels. Lessons progressed from segmenting words into syllables to segmenting words into phonemes, at first two phonemes and then three phonemes. Then blending was applied to the same words. Children performed manipulations with wooden markers at first and letters later on. Their work blending letters was the equivalent of learning to decode, and their work segmenting with letters was equivalent to learning to spell the sounds in words. More letters were added to the set later in the program. Words with consonant clusters were introduced. Finally two-syllable words were added. The program included various games, worksheets, and activities to teach these skills.

Teachers attended a half-day session to learn about the program, which was fully presented and described in a manual. The 17 teachers were asked to use the program 20 minutes daily. Their instruction was closely monitored. Although there were 12 units, only a few teachers got through the entire program in the 26-week period.

Williams evaluated the ABDs program again the following year, this time not with volunteer teachers but with 20 teachers who were mandated to use the program. They completed on average 6.6 units, about half the program. The treatment groups were compared to untreated control groups. The influence of PA instruction on students' ability to decode words and nonwords was measured at the end of training. Effect sizes were large, $\underline{d} = 1.05$ for the 1st year, and $\underline{d} = 0.97$ for the 2nd year. This indicates that the ABDs program was highly effective at teaching decoding skill to disabled readers.

Other Programs to Teach PA

Various programs were used to teach PA across studies. Presenting descriptions of these programs serves to clarify how studies in the database were structured and the variety of ways that PA was taught. Some programs had special features that enhanced their effectiveness. In the study by Cunningham (1990), one treatment group was taught metacognitive skills along with PA. Cunningham worked with normally progressing readers in kindergarten and 1st grade. A puppet was utilized to interact with children. PA training was limited to the oral mode, with no letter-sound instruction. Training was conducted in small groups for 10 weeks. Three treatments were compared. One treatment group received PA training in segmenting and blending phonemes. Another group received a somewhat abbreviated version of this training and spent the extra time in metacognitive activities that included learning about the goals and purposes of each PA manipulation, reviewing how that lesson related to previous lessons, and observing and practicing how to use the skill for reading. The control group spent equal time engaged in a story listening treament.

Results showed that at the end of PA training, the two treatment groups outperformed the control group on measures of PA and reading in both grades. In addition, 1st graders who had received both PA and metacognitive training achieved higher reading scores than 1st graders receiving only PA training. One possible reason why the advantage was limited to 1st grade is that 1st graders, but not kindergartners, were receiving formal reading instruction concurrently in their classrooms, so they had a chance to apply their PA knowledge on a daily basis. In fact, some 1st graders told the experimenter that they used what they had been taught to decode words in their classroom reading groups. These findings indicate that a metacognitive component may be valuable in providing a bridge between PA skills and reading processes. This may be particularly true in PA programs that do not teach phoneme manipulation with letters.

The ADD program (Auditory Discrimination in Depth) was developed by Lindamood and Lindamood (1975) to teach PA. The unique feature of this program is that it teaches children to identify and monitor articulatory gestures associated with phonemes. As already discussed, phoneme segmentation is difficult because there are no boundaries in speech telling us where one phoneme ends and the next begins. Rather phonemes are coarticulated to produce speech without any seams. One very helpful way to identify separate phonemes is to monitor the changes that occur in the mouth as one pronounces words. This involves directing attention to the position and shape of the lips and tongue. For example, there are three phonemes in meat and these are reflected in three successive mouth movements: your lips closing for /m/, your lips opening into a smile shape for the vowel, then your tongue tapping the roof of your mouth for /t/. Pictures of mouth positions can be used to help children distinguish phonemes in pronunciations of words. Also, mirrors help children explore what their own mouths are doing when they pronounce words.

Four studies in the NRP database implemented the ADD program to teach PA (Kennedy & Backman, 1993; McGuinness et al., 1995; (Wise, King, & Olson, 1999; Wise, King, & Olson, in press). Children received extensive training discovering and categorizing the various phonemes in English by analyzing their own mouth movements, often using mirrors. They learned to label these sounds, for example, lip poppers, tip tappers, and scrapers. They learned to track movements in spoken words in order to identify the separate phonemes and then to represent the phonemes with graphemes. Effect sizes on reading outcomes were variable, ranging from 1.22 for 1st graders (McGuinness et al., 1995) to 0.15 for older disabled readers (Wise, King, & Olson, 1999).

An example of a program focused on teaching only one type of phoneme manipulation was that studied by Byrne and Fielding-Barnsley (1991) for preschoolers, called Sound Foundations. This program taught phoneme identity. Children learned to recognize instances of the same sound in both initial and final positions across different words. The following sounds received primary attention: /s/, /š/ as in ship, /l/, /m/, /p/, /t/, /g/, /ae/ as in bat, /ɛ/ as in bet. Children were shown several large posters covered with pictures of objects. Their job was to pick out from a larger set the objects having a specified beginning or ending sound, for example, sea, seal, sailor, sand. Also, children were shown an array of pictures on worksheets or cards, and they selected those having targeted sounds. In each session, one phoneme in one position was taught. The letter representing that phoneme was introduced as well.

In this study, preschoolers averaging 4.5 years of age received either the PA training described above or control training that focused on story reading and semantic activities with the same posters and worksheets. Children were trained in groups of 4 to 6 children, one 30-minute lesson per week for 12 weeks. At the end of training, children in the PA-trained group were able to identify substantially more initial and final phonemes in words than control students. They demonstrated superior skill identifying not only sounds they had practiced but also unpracticed sounds, indicating that phoneme identity skill transferred to untaught phonemes. These researchers also gave students a simplified word reading task in which children were shown a word and identified it from two spoken choices (e.g., "Does this [sat] say sat or mat?"). Trained students read more words than control students, indicating that PA training improved preschoolers' rudimentary word recognition skill.

These researchers also investigated the long-term impact of PA training (Byrne & Fielding-Barnsley, 1993, 1995). Children were tested during the next 3 years in school. At the end of kindergarten, trained children were only slightly superior to controls in PA, indicating that learning to read had narrowed the gap in PA between the two groups. At the end of each successive grade, the PA-trained group read significantly more pseudowords than controls, indicating that PA training benefited children's decoding skill. At the end of 2nd grade, there was a marginal difference in reading comprehension favoring the PA-trained students. However, the 2nd graders did not differ in reading real words or in spelling words.

One possible reason why long-term training effects were not stronger in this study is that the formal reading and spelling instruction that children received in school was sufficiently effective to compensate for the advantage provided by preschool training in PA. Also, the PA training that students received was focused rather than comprehensive and amounted to only 6 hours total. It may take a more comprehensive and extensive training program to exert stronger long-term effects.

The effectiveness of different ways to teach PA was examined by O'Connor et al. (1995), who inquired whether PA training has to be broad rather than focused to be most effective. They selected at-risk kindergartners with low PA and randomly assigned them to one of three training conditions. In the comprehensive treatment, children performed a variety of sound manipulation activities that included isolating, segmenting, blending, and deleting phonemes; segmenting and blending syllables and onset-rime units; and working with rhyming words. In the focused treatment, children practiced segmenting and blending onsets, rimes, and phonemes only. Training extended for 10 weeks, two 15-minute sessions per week, totaling 5 hours. Beginning in the 5th week, letter-sound associations were taught for the sounds being practiced orally in both groups. However, children were not taught how to use letters to manipulate phonemes in the PA activities. The third treatment, a control condition, received only the letter-sound instruction.

Comparison of phonemic awareness following training showed that the treated groups performed equally well and both outperformed controls, indicating that both types of training were equally effective in teaching PA. To measure transfer to reading, a simplified word learning task was devised. After children learned to associate four letters and sounds, they were given practice learning to read five words composed of the letters and sounds: am, at, mat, sat, sam. Each word was taught by saying, "This is aaaaat, at." Results revealed that only the focused group learned to read the words in fewer trials than the control group, not the comprehensive group. This suggests that concentrating instructional time on segmenting and blending may contribute more to reading skill than diverting attention to many PA activities. These findings are consistent with those in the NRP meta-analysis indicating the greater impact of segmenting and blending than multiskill instruction on reading outcomes. One might question the use of a simplified word reading task to draw inferences about general reading acquisition. However, these kindergartners were beginning readers so that more advanced reading tests would have been too difficult.

The separate and combined contributions of instruction in segmentation and blending were examined by Davidson and Jenkins (1994), who gave kindergartners with low PA one or another of four types of training. In the segmentation treatment, each word was pronounced, and children were taught to say its separate sounds. In the blending treatment, children listened to the separate sounds and learned to blend them into words. In the segmentation-and-blending treatment, children learned first to segment, then to blend the words. In the control condition, children listened to stories. Children were taught to a criterion of mastery. The words and nonwords analyzed during training had two phonemes formed out of continuant consonants and long vowels (e.g., my, vo, low, way). At the end of training, all students were taught eight letters for the sounds that treatment groups had practiced. Then two literacy tests were given in which children practiced and received feedback in learning to read and learning to spell two-phoneme words. These words were formed from the same letter-sounds but they had not been taught during training.

Results showed that the groups learned the PA skill that they were taught but performed poorly on the untaught skill. This indicates that teaching students either segmentation or blending does not improve their performance in the other skill. On the measures of reading and spelling, both the segmentation and combination groups performed similarly and outperformed the control group. However, the blending group did not do better than the control group. This indicates that teaching beginners to segment is as effective for learning to read words as teaching beginners to segment and blend. In contrast, teaching beginners only to blend is not effective. These findings were replicated in a similar study by Torgesen et al. (1992).

Although blending made a poor showing in these studies, Reitsma and Wesseling (1998) reported more success in a study with kindergartners in the Netherlands. They used a computer to teach kindergartners how to blend three-phoneme Dutch words (e.g., lief, geit, met). No limits were placed on the variety of phonemes in the words. All phoneme manipulations were conducted in speech without any letters. First, children were taught a set of vocabulary words, and then these were used in various blending exercises. Children listened to a sequence of segmented sounds, and then clicked on the picture corresponding to that word. Children listened to two successively segmented words and clicked "same" or "different." Children listened to words, either pronounced as wholes or segmented, and then had to find which of several boxes on the screen contained the other form of the word. If a whole word was heard, they had to find its segmented form. If a segmented word was heard, they had to find its whole form. In all these exercises, the incorrect word choices differed by several phonemes from the correct choice for some items but only by one phoneme for other items, making processing more difficult. In the control group, children completed vocabulary exercises on the computer.

At the end of kindergarten, PA tests of children's ability to blend and to segment words revealed superior blending performance by the trained group over the control group, but no difference in segmentation performance. Thus, training effects were limited to blending which was the skill taught, and blending skill did not transfer to segmentation. The following year, in 1st grade, children's ability to read words was examined. Long-term effects of the blending exercises were evident because trained children read more words than control children. However, no effects on spelling were detected. These results suggest that extensive training to develop blending skills does benefit reading acquisition. Blending is thought to contribute to reading by enabling children to decode new words they have not yet learned to read. Also, findings indicate the effectiveness of using computers to teach PA to kindergartners.

One instructional activity that is maximally effective for teaching PA in a way that builds a bridge to reading and spelling is that of teaching children to invent phonemically more complete spellings of words. Typically, kindergartners who know letter names or sounds can represent the more salient sounds in words such as beginning and ending sounds, for example, writing B to spell beaver or R to spell arm. Sometimes their spellings are not conventional, for example, writing Y to spell wife. However, the important achievement is that they can distinguish sounds in words. Once they can do this, then teachers can help them detect additional sounds in words and learn conventional spellings for those sounds.

In a study by Ehri and Wilce (1987b), kindergartners were taught individually how to generate phonemic spellings of words and nonwords by segmenting words into phonemes and selecting letters representing those phonemes. Children who qualified for the study could already name the six consonant and four vowel letters that were used in training. All names contained the relevant sounds in their names (T, S, N, L, K, P, A, E, I, O).

Instruction began with two-phoneme words and nonwords and progressed to three-phoneme words and words with consonant clusters. Children were helped to break words into phonemes by directing their attention to articulatory gestures. They were helped to select letters by focusing on sounds in letter names. They mastered shorter words before advancing to longer words. Children in the control group practiced matching the ten letters and sounds in isolation. Articulatory gestures and letter names were used to correct their errors as well. On posttests after training, effect sizes were large on measures of segmentation and spelling. The measure of reading involved giving children practice learning to read 12 similarly spelled words for several trials. The words were spelled phonemically with the letter-sounds taught, for example, SEL (seal), SNAK (snake), SLIS (slice). The effect size was large, $d = 0.97$. These findings indicate that teaching children to segment and spell helps them learn to read as well as spell words.

In many PA training studies, the instructional context was not considered. However, there were some exceptions. Iversen and Tunmer (1993) incorporated PA training into Clay's (1985) Reading Recovery program to examine whether systematic instruction in PA would make the program more effective. At-risk readers in 1st grade were assigned to one of three groups, a group receiving standard Reading Recovery instruction, a group receiving modified RR instruction, and an alternative, non-RR intervention group. In the modified RR treatment, after children had learned most letters, they manipulated magnetic letter forms to make, break, and build new words having similar spellings and pronunciations, for example, reading and and then changing it to hand, sand, band. Training progressed from initial sounds to final sounds and then to medial sounds. Children added, deleted, and substituted letters in their manipulations and also read the changed words. Later, the task becomes a writing rather than a manipulation task.

Findings showed that both forms of RR enabled children to reach prescribed reading levels that qualified them to exit the remedial program. However, children who received modified RR attained prescribed levels more quickly than children receiving the standard program (i.e., a mean of 41.75 lessons for modified RR vs. 57.31 lessons for standard RR). This indicates that adding PA training improved RR by increasing its efficiency. At the end of training, however, both groups performed at very similar levels on PA outcomes and reading outcomes, indicating that both forms of the RR program enabled children to attain similar levels of PA and reading. On followup tests given at the end of the school year, performance of the groups remained very similar.

Hatcher et al. (1994) also examined whether adding PA training to a Reading Recovery® program would improve its success. The participants were 7-year-old poor readers. The PA training that was added to RR involved teaching children to perform different types of PA, including segmentation, blending, deletion, substitution, and transposition of phonemes. Children also practiced linking letters to phonemes in various spelling and writing tasks. Effect sizes, though small, favored the PA-trained group (*d* = 0.24 for PA, *d* = 0.31 for reading and spelling).

Castle et al. (1994) examined the contribution of PA training to reading acquisition in a whole language program. Kindergartners with low PA were assigned to treatment and control groups. PA training included segmentation, blending, substitution, and deletion. Letters were incorporated into the PA activities later in the program. Two control groups were included, one receiving an alternative, unrelated treatment and another receiving no treatment other than the whole language instruction provided to all participants in their classrooms. Results showed that the PA-trained group spelled more words and decoded many more pseudowords than the two control groups. However, the groups did not differ in reading real words or in reading connected text. These findings indicate that adding PA instruction to a whole language program enhances students' decoding and spelling skills but not their other reading skills.

Wise et al. (in press) evaluated the effects of PA training against training that taught children reading comprehension strategies and gave them extensive text reading practice on computers. The children were 200 disabled readers in grades 2 to 5. Both treatment and control groups spent time reading stories on the computer. They could touch any unknown word with a cursor and have it identified. Comprehension questions were answered periodically. Controls spent extra time reading on the computer while the PA-trained group completed various types of PA activities administered by the computer. For example, the computer asked the child to show *feef*. The child selected and ordered letter-sound symbols with a mouse. Synthetic speech pronounced whatever the child assembled, and the child continued to manipulate letters until achieving a match. Then the computer asked the child to change the word to *feem*. Lessons began with two-phoneme words and progressed to longer words. There were several other PA activities besides this one.

On the posttests, PA-trained children outperformed controls on tests of phonemic awareness and pseudoword reading tests. Also, they read more words when there were no time constraints. However, controls displayed superior time-limited word reading. Both groups made similar gains in spelling and reading comprehension. Interestingly, when the analysis of word reading took account of grade level, 2nd graders gained more than older children and they showed a much greater advantage for PA training over the control training than did older children. These findings suggest that PA training may be more beneficial to younger than to older disabled readers.

In sum, these illustrative studies enrich the understanding of the data contributing to the NRP meta-analysis. They show that various types of instruction were utilized to teach PA at various grade levels. They show how different studies were designed and the nature of their findings. Also, they draw attention to other potentially important features that were not addressed in the meta-analysis because of an inadequate number of cases.

Implications for Reading Instruction

1. Can phonemic awareness be taught, and does it help children learn to read and spell?

Results of the meta-analysis showed that teaching phonemic awareness to children is clearly effective. It improves their ability to manipulate phonemes in speech. This skill transfers and helps them learn to read and spell. PA training benefits not only word reading but also reading comprehension. PA training contributes to children's ability to read and spell for months, if not years, after the training has ended. Effects of PA training are enhanced when children are taught how to apply PA skills to reading and writing tasks.

2. Which students benefit in their reading?

Teaching phonemic awareness helps many different students learn to read, including preschoolers, kindergartners, and 1st graders who are just starting to learn to read. This includes beginners who are low in PA and are thus at risk for developing reading problems in the future. This includes older disabled readers who have already developed reading problems. This includes children from various SES levels. This includes students who are taught to read in English, as well as students taught to read in other alphabetic languages.

3. Which students benefit in their spelling?

Teaching phonemic awareness helps preschoolers, kindergartners, and 1st graders learn to spell. It helps children at risk for future reading problems also. It helps low as well as middle-to-high SES children. It helps students learning to spell in English as well as students learning in other languages. However, PA training is ineffective for improving spelling in reading-disabled students. This is consistent with other research indicating that disabled readers have a hard time learning to spell.

4. Which methods of teaching PA work best in helping children acquire phonemic awareness?

Various forms of phoneme manipulation might be taught, including identifying or categorizing the phonemes in words, segmenting words into phonemes, blending phonemes to form words, deleting phonemes from words, or manipulating onsets and rimes in words. In some programs, only one PA skill is taught, while in other programs, two or more skills are combined. Some programs teach children to use letters to manipulate phonemes and others limit training to speech. All of these approaches appear to be effective for helping children learn to manipulate phonemes. Focusing on one or two skills produces larger effects than a multiskilled approach. Teaching PA with letters helps students acquire PA more effectively than teaching PA without letters.

5. Which methods of teaching PA have the greatest impact on learning to read?

Although all of the approaches exert a significant effect on reading, instruction that focuses on one or two skills produces greater transfer than a multiskilled approach. Teaching students to segment and blend benefits reading more than a multiskilled approach. Teaching students to manipulate phonemes with letters yields larger effects than teaching students without letters, not surprisingly because letters help children make the connection between PA and its application to reading. Teaching children to blend the phonemes represented by letters is the equivalent of decoding instruction. Being explicit about the connection between PA skills and reading also strengthens training effects.

6. Which methods of teaching PA have the greatest impact on learning to spell?

Teaching PA helps nondisabled readers below 2nd grade learn to spell. Methods that teach children to manipulate phonemes with letters are more effective than methods limiting manipulation to spoken units. Teaching children to segment phonemes in words and represent them with letters is the equivalent of invented spelling instruction.

7. How important is it to teach letters as well as phonemic awareness?

It is essential to teach letters as well as phonemic awareness to beginners. PA training is more effective when children are taught to use letters to manipulate phonemes. This is because knowledge of letters is essential for transfer to reading and spelling. Learning all the letters of the alphabet is not easy, particularly for children who come to school knowing few of them. Shapes, names, and sounds need to be overlearned so that children can work with them automatically to read and spell words. Thus, if children do not know letters, this needs to be taught along with PA.

8. How much time is required for PA instruction to be effective?

In the NRP analysis, studies that spent between 5 and 18 hours teaching PA yielded very large effects on the acquisition of phonemic awareness. Studies that spent longer or less time than this also yielded significant effect sizes, but effects were moderate and only half as large. Transfer to reading was greatest for studies lasting less than 20 hours. In fact, effect sizes were more than twice as large for shorter programs than for the longest-lasting programs.

Caution is needed in drawing conclusions from this finding. Although it suggests that less instructional time is better, it ignores reasons why training that lasted longer might have been less effective. Perhaps the PA skills being taught were more complex, or perhaps the learners were harder to teach, or perhaps, as a result of time spent in training, PA-trained students received less instruction in reading than students in the control groups.

The Panel concludes that it is wrong to make any declarations about how long effective instruction in PA needs to last based on the NRP findings. Rather, decisions should be influenced by reason, moderation, and situational factors. The answer depends on the goals of instruction, how many different PA skills are to be taught, whether letters are included, how much or how little the learners already know about PA when they begin, whether they are disabled readers, whether provision is made for facilitating transfer to reading and spelling, and so forth. Individual children will differ in the amount of training time they need to acquire PA. What is probably most important is to tailor training time to student learning by assessing who has and who has not acquired the skills being taught as training proceeds. Children who are still having trouble should continue PA training while those who have learned the skills should move on to other reading and writing instruction.

Not only the total training time but also the length of single training sessions must be considered. In the NRP database, the average length of sessions was 25 minutes. Few sessions lasted more than 30 minutes, and these tended to occur with older disabled readers, not with younger children. From this, the Panel concludes that sessions should probably not exceed 30 minutes in length.

9. Can classroom teachers teach PA effectively to their students?

Classroom teachers are definitely able to teach PA effectively. In the NRP analysis, their effect size on the acquisition of PA was large. The training they provided transferred and improved students' reading and spelling, and the effect on reading continued beyond training. It was not possible to specify the amount of training required to enable trainers to be effective. This relationship was not examined in the studies. Only 15 studies reported the length of training provided to trainers. It ranged from 2 to 90 hours, with a mean of 21 hours. This suggests that the amount of training required may be quite modest and reasonable for inservice instruction.

10. Is instruction most effectively delivered to individual students, to small groups, or to full classrooms of students?

Although individual tutoring is commonly regarded as the most effective unit of instruction, NRP findings indicate that small groups are the best way to teach phonemic awareness to children. Also, small groups facilitate greater transfer to reading than the other two teaching units. This may hold true for several reasons. Children may benefit from observing their peers respond and receive feedback or from listening to their peers' comments and explanations. Or children may be more attentive and motivated to learn so that they do well in the eyes of their peers.

11. Is evidence for the effectiveness of PA training on reading outcomes derived from strongly designed or weakly designed studies?

The NRP analyses show that the evidence rests solidly on well-designed studies. Significant effect sizes were apparent on standardized tests as well as experimenter-designed tests. Random assignment of children to groups yielded significant effects. In fact, this effect size was larger than that for the nonequivalent group design. Studies in which treatment fidelity was checked yielded a moderate effect size. Significant effects occurred not only when PA-trained groups were compared to untreated control groups but also when they were compared to treated controls. Significant effects were detected with larger as well as smaller samples of children. When Troia's (1999) criteria for methodological rigor were applied to studies, the most rigorous studies yielded the largest effect sizes. The Panel concludes that evidence for the effectiveness of PA training on reading outcomes comes from well-designed experiments. In fact, researchers are advised that they have the best chance of observing strong effects if they apply the most rigor in designing their PA studies.

12. Are the results ready for implementation in the classroom?

This section of the NRP report includes many ideas that provide guidance to teachers in designing PA instruction and in evaluating and selecting programs with the best chance for success. However, in implementing PA instruction in the classroom, teachers should bear in mind several serious cautions:

- PA training does not constitute a complete reading program. Although the present meta-analysis confirms that PA is a key component that contributes significantly to the effectiveness of beginning reading and spelling instruction, there is obviously much more that children need to be taught to acquire reading and writing competence. PA instruction is intended only as a foundational piece. It helps children grasp how the alphabetic system works. It helps children read and spell words in various ways. However, literacy acquisition is a complex process for which there is no single key to success. Teaching phonemic awareness does not ensure that children will learn to read and write. Many competencies must be acquired for this to happen.
- Exactly how PA instruction should be taught by teachers in their classrooms is not clearly specified by the research. A variety of programs was found to be effective. The studies are useful in identifying features that are important and should be considered in selecting programs and planning classroom instruction. Ultimately, though, teachers need to evaluate the methods they use against measured success in their own students.
- One factor that is very important to effective classroom instruction but has not been addressed in the PA training research is the extent to which these programs motivate both students and teachers. It seems self-evident that instructional techniques for developing PA need to be relevant, engaging, interesting, and motivating in order to promote optimal learning in children. However, the research has not focused on this factor. Neither has the research examined which techniques are most engaging for teachers. It seems self-evident that teachers are most effective when they are enthusiastic and enjoy what they are teaching. In selecting ways to teach PA, teachers need to take account of motivational aspects of programs for themselves as well as their students.
- Teachers should recognize that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping learners understand and use the alphabetic system to read and write. This is why it is important to include letters when teaching children to manipulate phonemes and why it is important to be explicit about how children are to use the PA skills in reading and writing tasks.
- It is important to recognize that children will acquire some phonemic awareness in the course of learning to read and spell even though they are not taught PA explicitly. The process of learning letter-sound relations and how to use them to read and spell enhances children's ability to manipulate phonemes. However, incidental instruction that does not focus on teaching PA falls short in its contribution to children's reading and spelling development.
- It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness; so, PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling already while others may know only a few letters and have no reading skill. The nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA prior to beginning PA instruction. This will indicate which children need the instruction and which do not; which children need to be taught rudimentary levels of PA, for example, segmenting initial sounds in words; and which need more advanced levels involving segmenting or blending with letters.

Directions for Further Research

A large number of experiments have been conducted to test whether phonemic awareness training helps children learn to read. Results have been sufficiently positive to sustain confidence that this treatment is indeed effective across a variety of child and training conditions. However, there are still some questions needing further attention from researchers.

1. Training Teachers to Teach PA

Findings of a few studies have raised doubt that teachers possess sufficient phonemic awareness to teach this skill adequately on their own (Moats, 1994; Scarborough, Ehri, Olson, & Fowler, 1998). These studies indicate that teachers fall short in manipulating phonemes correctly. However, the studies do not show that this lack of knowledge limits teachers' ability to *learn* to teach PA adequately. Results of the Panel's analysis indicate that with training, teachers can teach PA effectively.

Research is needed to clarify what sort of knowledge and training maximizes teachers' effectiveness in teaching PA and in integrating this instruction with beginning reading instruction. This includes both preservice training and inservice training that covers instruction for preschoolers, primary students, and older disabled readers. Questions to be addressed are: How much and what sort of linguistic knowledge about phonemes, graphemes, and the alphabetic system need to be taught to teachers? How much knowledge about literacy learning processes and their course of development in beginning readers needs to be understood by teachers? Teachers may need to know how phonemic awareness develops in children, which tasks are easier and which are harder, what techniques help children focus on phoneme-size units such as monitoring articulatory cues, what kinds of mistakes children commonly make, what the origin is of these mistakes, how they should be corrected, and so forth. Teaching children to invent spellings of words is one way to teach PA. Teachers may need to understand the processes children use to invent spellings, how their spellings become more complete and conventional, and how to promote this growth. Such knowledge should help teachers utilize this approach to teach PA. Research is needed to address these possibilities.

2. Use of Small Groups, Large Groups, or Individual Tutoring to Teach PA

In the meta-analysis of instructional programs, size of training unit was uncovered as a property that affected outcomes differentially. Small group instruction was associated with much larger effect sizes than individual or classroom instruction. However, these findings are correlational. That is, differences emerged across studies. Differences did not arise in studies that manipulated this variable experimentally. As a result, attributing cause to this property is highly tentative and open to other interpretations. The next step for researchers is to determine experimentally whether small group instruction is indeed a better way to teach PA than individual and classroom instruction and, if so, the processes and conditions that make this approach especially effective.

3. Motivation to Teach and to Learn PA

Research has focused on the cognitive and linguistic factors involved in teaching PA to children. However, if teachers are not motivated to teach this skill, or if children are not motivated to learn it, then attention to it may be slighted. Some forms of teaching and learning are interesting and fun whereas other forms are tedious and boring. Research is needed to assess motivational properties of PA training programs and ways of enhancing motivation and interest if they are lacking.

4. Teaching PA With Computers

Use of computers is fast becoming a national pastime at home as well as at school. Younger children are acquiring facility with computers. Parents, as well as teachers are in the market for effective computer programs to teach important skills to children. A few studies in the NRP database examined whether computers could deliver PA instruction effectively. Findings showed that effect sizes were significant for teaching PA and its transfer to reading. However, effects were smaller than those produced by teachers or researchers. Computers were of doubtful value for promoting transfer to spelling although this may apply only to older disabled readers. More research is needed to determine whether and how PA might be taught more effectively using computers.

5. Programs to Help Parents Teach PA

Many parents of preschoolers are anxious to help their children acquire the knowledge and skills they need to become successful when they enter school and begin reading instruction. However, none of the studies reviewed utilized parents as trainers. Research is needed to address this gap in our knowledge. In addition to informal activities that parents might use to draw children's attention to sounds in words, the effectiveness of activities that help parents teach letters to preschoolers might be explored and assessed.

6. High-Quality Research

Results of the NRP meta-analysis reveal the value of experimental studies for providing reliable findings that can guide instructional practice. The Panel examined whether well-designed studies yielded stronger effect sizes than weaker designs and found that effect sizes were largest for studies that were methodologically rigorous. It is important for future researchers to maintain the quality of the designs adopted. This is not to say that all studies must use random assignment rather than nonequivalent groups. Sometimes experimenters have no choice if they want to conduct studies in school classrooms. However, researchers must take steps to maximize the rigor of their studies by addressing as many threats to internal and external validity as possible. Not only does this enhance confidence in the findings but also, as the NRP meta-analysis shows, it gives researchers a better chance of detecting treatment effects when they exist.

References

Note: An asterisk marks those publications providing data for the meta-analysis.

Adams, M. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Adams, M., Foorman, B., Lundberg, I., & Beeler, T. (1998). Phonemic awareness in young children: A classroom curriculum. Baltimore, MD: Brookes Publishing Co.

Anderson, R., Hiebert, E., Scott, J. & Wilkinson, I. (1985). Becoming a nation of readers: The report of the commission on reading. Washington, DC: National Academy of Education, Commission on Education and Public Policy.

Apthorp, H. (1998, April). Phonological awareness training for beginning readers: A meta-analysis of reading posttests. Paper presented at the American Educational Research Association Annual Meeting, San Diego, CA.

* Ball, E., & Blachman, B. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? Reading Research Quarterly, 26, 49-66.

Balmuth, M. (1982). The roots of phonics: a historical introduction. New York: Teachers College Press.

* Barker, T., & Torgesen, J. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. Journal of Educational Computing Research, 13, 89-103.

* Bentin, S., & Leshem, H. (1993). On the interaction between phonological awareness and reading acquisition: It's a two-way street. Annals of Dyslexia, 43, 125-148.

Blachman, B. (in press). Phonological awareness. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), Handbook of reading research (Vol. 3). Mahwah, NJ: Erlbaum.

* Blachman, B., Ball, E., Black, R., & Tangel, D. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? Reading and Writing: An Interdisciplinary Journal, 6, 1-18.

Bloom, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13, 4-16.

* Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. Nature, 301, 419-421.

* Bradley, L., & Bryant, P. (1985). Rhyme and reason in reading and spelling. International Academy for Research in Learning Disabilities, Monograph Series, 1, 75-95. Ann Arbor, MI: The University of Michigan Press. (A more complete report of Bradley & Bryant, 1983)

* Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. Annals of Dyslexia, 44, 26-59.

* Brennan, F., & Ireson, J. (1997). Training phonological awareness: A study to evaluate the effects of a program of metalinguistic games in kindergarten. Reading and Writing: An Interdisciplinary Journal, 9, 241-263.

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. Developmental Psychology, 28, 874-886.

Bruck, M. (1993). Component spelling skills of college students with childhood diagnoses of dyslexia. Learning Disability Quarterly, 16, 171-184.

* Bus, A. (1986). Preparatory reading instruction in kindergarten: Some comparative research into methods of auditory and auditory-visual training of phonemic analysis and blending. Perceptual and Motor Skills, 62, 11-24.

Bus, A., & van Ijzendoorn, M. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. Journal of Educational Psychology, 91, 403-414.

* Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. Journal of Educational Psychology, 83, 451-455.

* Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. Journal of Educational Psychology, 85, 104-111. (First followup to Byrne & Fielding-Barnsley, 1991)

* Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. Journal of Educational Psychology, 87, 488-503. (second followup to Byrne & Fielding-Barnsley, 1991)

* Castle, J. M., Riach, J., & Nicholson, T. (1994). Getting off to a better start in reading and spelling: The effects of phonemic awareness instruction within a whole language program. Journal of Educational Psychology, 86, 350-359.

Chall, J. (1967). Learning to read: the great debate. New York: McGraw Hill.

Clay, M. (1985). The early detection of reading difficulties (3rd ed.). Auckland, New Zealand: Heinemann.

Cohen, J. (1988). Statistical power analysis for the behavior sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, P., Kulik, J., & Kulik, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. American Educational Research Journal, 19, 237-248.

Cooper, H. (1998). Synthesizing research. Thousand Oaks, CA: Sage.

* Cunningham, A. (1990). Explicit versus implicit instruction in phonemic awareness. Journal of Experimental Child Psychology, 50, 429-444.

* Davidson, M., & Jenkins, J. (1994). Effects of phonemic processes on word reading and spelling. Journal of Educational Research, 87, 148-157.

* Defior, S., & Tudela, P. (1994). Effect of phonological training on reading and writing acquisition. Reading and Writing: An Interdisciplinary Journal, 6, 299-320.

Dressman, M. (1999). On the use and misuse of research evidence: Decoding two states' reading initiatives. Reading Research Quarterly, 34, 258-285.

Ehri, L. (1979). Linguistic insight: Threshold of reading acquisition. In T. G. Waller & G. E. MacKinnon (Eds.), Reading research: Advances in theory and practice (Vol. 1, pp. 63-114). New York: Academic Press.

Ehri, L. (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), Language awareness and learning to read (pp. 119-147). New York: Springer Verlag.

Ehri, L. (1991). Development of the ability to read words. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research, (Vol. 2, pp. 383-417). New York: Longman.

Ehri, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), Reading acquisition (pp. 107-143). Hillsdale, NJ: Erlbaum.

Ehri, L. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed., pp. 323-358). Newark, DE: International Reading Association.

Ehri, L., & Wilce, L. (1987a). Cipher versus cue reading: An experiment in decoding acquisition. Journal of Educational Psychology, 79, 3-13.

* Ehri, L., & Wilce, L. (1987b). Does learning to spell help beginners learn to read words? Reading Research Quarterly, 22, 48-65.

* Farmer, A., Nixon, M., & White, R. (1976). Sound blending and learning to read: An experimental investigation. British Journal of Educational Psychology, 46, 155-163.

Fawcett, A., & Nicolson, R. (1995). Persistence of phonological awareness deficits in older children with dyslexia. Reading and Writing: An Interdisciplinary Journal, 7, 361-376.

* Fox, B., & Routh, D. (1976). Phonemic analysis and synthesis as word-attack skills. Journal of Educational Psychology, 68, 70-74.

* Fox, B., & Routh, D. (1984). Phonemic analysis and synthesis as word-attack skills: Revisited. Journal of Educational Psychology, 76, 1059-1064.

Gaskins, I., Downer, M., Anderson, R., Cunningham, P., Gaskins, R., Schommer, M., & The Teachers of Benchmark School. (1988). A metacognitive approach to phonics: Using what you know to decode what you don't know. Remedial and Special Education, 9, 36-41.

Glass, G., Cahen, L., Smith, M., & Filby, N. (1982). School class size. Beverly Hills, CA: Sage.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5, 674-691.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. Journal of Experimental Child Psychology, 42, 73-83.

Griffith, P. (1991). Phonemic awareness helps first graders invent spellings and third graders remember correct spellings. Journal of Reading Behavior, 23, 215-233.

* Gross, J., & Garnett, J. (1994). Preventing reading difficulties: Rhyme and alliteration in the real world. Educational Psychology in Practice, 9, 235-240.

* Haddock, M. (1976). Effects of an auditory and an auditory-visual method of blending instruction on the ability of prereaders to decode synthetic words. Journal of Educational Psychology, 68, 825-831.

* Hatcher, P., Hulme, C., & Ellis, A. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. Child Development, 65, 41-57.

* Hohn, W., & Ehri, L. (1983). Do alphabet letters help prereaders acquire phonemic segmentation skill? Journal of Educational Psychology, 75, 752-762.

* Hurford, D., Johnston, M., Nepote, P., Hampton, S., Moore, S., Neal, J., Mueller, A., McGeorge, K., Huff, L., Awad, A., Tatro, C., Juliano, C., & Huffman, D. (1994). Early identification and remediation of phonological-processing deficits in first-grade children at risk for reading disabilities. Journal of Learning Disabilities, 27, 647-659.

* Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery Program. Journal of Educational Psychology, 85, 112-126.

Johnson, B. (1989). DSTAT: Software for the meta-analytic review of research literatures. Hillsdale, NJ: Erlbaum.

Johnson, B., & Eagly, A. (in press). Quantitative synthesis of social psychological research. In H. Reis & C. Judd (Eds.), Handbook of research methods in social psychology (pp. 496-528). London: Cambridge University Press.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. Journal of Educational Psychology, 80, 437-447.

Juel, C., Griffith, P., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. Journal of Educational Psychology, 78, 243-255.

* Kennedy, K., & Backman, J. (1993). Effectiveness of the Lindamood Auditory Discrimination in depth program with students with learning disabilities. Learning Disabilities Research and Practice, 8, 253-259.

* Korkman, M., & Peltomaa, A. (1993). Preventive treatment of dyslexia by a preschool training program for children with language impairments. Journal of Clinical Child Psychology, 22, 277-287.

* Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. Learning and Instruction, 5, 187-201.

Liberman, I., Shankweiler, D., Fischer, F., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

* Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. Reading Research Quarterly, 26, 234-250.

Lindamood, C. & Lindamood, P. (1975). The ADD program: Auditory discrimination in depth. Austin, TX: DLM.

* Lovett, M., Barron, R., Forbes, J., Cuksts, B., & Steinbach, K. (1994). Computer speech-based training of literacy skills in neurologically impaired children: A controlled evaluation. Brain and Language, 47, 117-154.

* Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. Reading Research Quarterly, 23, 263-284.

Mann, V. (1987). Phonological awareness: The role of reading experience. In P. Bertelson (Ed.), The onset of literacy: Cognitive processes in reading acquisition (pp. 65-92). Cambridge, MA: The MIT Press.

Marsh, G., Freidman, M., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. In G. Mackinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3, pp. 199-221). New York: Academic Press.

* McGuiness, D., McGuiness, C., & Donohue, J. (1995). Phonological training and the alphabet principle: Evidence for reciprocal causality. Reading Research Quarterly, 30, 830-852.

Moats, L. (1994). Knowledge of language: The missing foundation for teacher education. Annals of Dyslexia, 44, 81-102.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1987). Literacy training and speech segmentation. In P. Bertelson (Ed.), The onset of literacy: cognitive processes in reading acquisition (pp. 45-64). Cambridge, MA: The MIT Press.

* Murray, B. (1998). Gaining alphabetic insight: Is phoneme manipulation skill or identity knowledge causal? Journal of Educational Psychology, 90, 461-475.

* O'Connor, R., & Jenkins, J. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. The Journal of Special Education, 29, 255-275.

* O'Connor, R., Jenkins, J., & Slocum, T. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. Journal of Educational Psychology, 87, 202-217.

* O'Connor, R., Notari-Syverson, A., & Vadasy, P. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. Exceptional Children, 63, 117-130.

* O'Connor, R., Notari-Syverson, A., & Vadasy, P. (1998). First-grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. Learning Disabilities Research and Practice, 13, 43-52. (followup to O'Connor et al., 1996)

* Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? Scandinavian Journal of Psychology, 24, 35-44.

* Olofsson, A., & Lundberg, I. (1985). Evaluation of long-term effects of phonemic awareness training in kindergarten: Illustrations of some methodological problems in evaluation research. Scandinavian Journal of Psychology, 26, 21-34. (followup to Olofsson & Lundberg, 1983)

Pinnel, G., Lyons, C., DeFord, D., Bryk, A., & Seltzer M. (1994). Comparing instructional models for the literacy education of high-risk 1st graders. Reading Research Quarterly, 29, 9-39.

Rack, J., Hulme, C., Snowling, M., & Wightman, J. (1994). The role of phonology in young children learning to read words: The direct-mapping hypothesis. Journal of Experimental Child Psychology, 57, 42-71.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1987). The ability to manipulate speech sounds depends on knowing alphabetic writing. In P. Bertelson (Ed.), The Onset of Literacy: Cognitive Processes in Reading Acquisition (pp. 31-44). Cambridge, MA: The MIT Press.

Reitsma, P. (1983). Printed word learning in beginning readers. Journal of Experimental Child Psychology, 75, 321-339.

* Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergartners. Scientific Studies of Reading, 2, 301-320.

* Sanchez, E., & Rueda, M. (1991). Segmental awareness and dyslexia: Is it possible to learn to segment well and yet continue to read and write poorly? Reading and Writing: An Interdisciplinary Journal, 3, 11-18.

Scarborough, H., Ehri, L., Olson, R., & Fowler, A. (1998). The fate of phonemic awareness beyond the elementary school years. Scientific Studies of Reading, 2, 115-142.

Schatschneider, C., Francis, D., Foorman, B., Fletcher, J., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. Journal of Educational Psychology, 91, 439-449.

* Schneider, W., Kuspert, P., Roth, E., Vise, M., & Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. Journal of Experimental Child Psychology, 66, 311-340.

Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. Journal of Educational Psychology, 76, 1309-1324.

Snow, C., Burns, M., & Griffin, P. (Eds.), (1998). Preventing reading difficulties in young children. Washington DC: National Academy Press.

* Solity, J. (1996). Phonological awareness: Learning disabilities revisited? Educational and Child Psychology, 13, 103-113.

Stahl, S., & Murray, B. (1994). Defining phonological awareness and its relationship to early reading. Journal of Educational Psychology, 86, 221-234.

* Tangel, D., & Blachman, B. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. Journal of Reading Behavior, 24, 233-261.

* Torgesen, J., Morgan, S., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. Journal of Educational Psychology, 84, 364-370.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. Journal of Experimental Child Psychology, 39, 161-181.

* Treiman, R., & Baron, J. (1983). Phonemic-analysis training helps children benefit from spelling sound rules. Memory and Cognition, 11, 382-389.

Troia, G. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. Reading Research Quarterly, 34, 28-52.

Tunmer, W., & Chapman, J. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme and M. Joshi (Eds.), Reading and spelling: Development and disorders (pp. 33-68). Mahwah, NJ: Erlbaum.

* Uhry, J., & Shepherd, M. (1993). Segmentation/spelling instruction as part of a first-grade

reading program: Effects on several measures of reading. Reading Research Quarterly, 28, 218-233.

* Vadasy, P., Jenkins, J., Antil, L., Wayne, S., & O'Connor, R. (1997). Community-based early reading intervention for at-risk first graders. Learning Disabilities Research and Practice, 12, 29-39.

* Vadasy, P., Jenkins, J., Antil, L., Wayne, S., & O'Connor, R. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. Learning Disability Quarterly, 20, 126-139.

* Vellutino, F., & Scanlon, D. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. Merrill-Palmer Quarterly, 33, 321-363.

Venezky, R. (1970). The structure of English orthography. The Hague: Mouton.

Venezky, R. (1999). The American way of spelling. New York: Guilford Press.

Wagner, R. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. Merrill-Palmer Quarterly, 32, 261-279.

Wagner, R., & Torgesen, J. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological Bulletin, 101, 192-212.

* Warrick, N., Rubin, H., & Rowe-Walsh, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. Annals of Dyslexia, 43, 153-173.

Wasik, B., & Slavin, R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. Reading Research Quarterly, 28, 179-200.

* Weiner, S. (1994). Effects of phonemic awareness training on low- and middle-achieving first graders' phonemic awareness and reading ability. Journal of Reading Behavior, 26, 277-300.

* Williams, J. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. Journal of Educational Psychology, 72, 1-15.

* Wilson, J., & Frederickson, N. (1995). Phonological awareness training: An evaluation. Educational & Child Psychology, 12, 68-79.

* Wise, B., King, J., & Olson, R. (1999). Training phonological awareness with and without explicit attention to articulation. Journal of Experimental Child Psychology, 72, 271-304.

* Wise, B., King, J., & Olson, R. (in press). Individual differences in benefits from computer-assisted remedial reading. Journal of Experimental Child Psychology.

Appendix A

Studies Included in the Meta-Analysis

* Ball, E., & Blachman, B. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? Reading Research Quarterly, 26, 49-66.

* Barker, T., & Torgesen, J. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. Journal of Educational Computing Research, 13, 89-103.

* Bentin, S., & Leshem, H. (1993). On the interaction between phonological awareness and reading acquisition: It's a two-way street. Annals of Dyslexia, 43, 125-148.

* Blachman, B., Ball, E., Black, R., & Tangel, D. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? Reading and Writing: An Interdisciplinary Journal, 6, 1-18.

* Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. Nature, 301, 419-421.

* Bradley, L., & Bryant, P. (1985). Rhyme and reason in reading and spelling. International Academy for Research in Learning Disabilities, Monograph Series, 1, 75-95. Ann Arbor, MI: The University of Michigan Press. (a more complete report of Bradley & Bryant, 1983)

* Brady, S., Fowler, A., Stone, B., & Winbury, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. Annals of Dyslexia, 44, 26-59.

* Brennan, F., & Ireson, J. (1997). Training phonological awareness: A study to evaluate the effects of a program of metalinguistic games in kindergarten. Reading and Writing: An Interdisciplinary Journal, 9, 241-263.

* Bus, A. (1986). Preparatory reading instruction in kindergarten: Some comparative research into methods of auditory and auditory-visual training of phonemic analysis and blending. Perceptual and Motor Skills, 62, 11-24.

* Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. Journal of Educational Psychology, 83, 451-455.

* Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. Journal of Educational Psychology, 85, 104-111. (First followup to Byrne & Fielding-Barnsley, 1991)

* Byrne, B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. Journal of Educational Psychology, 87, 488-503. (Second followup to Byrne & Fielding-Barnsley, 1991)

* Castle, J. M., Riach, J., & Nicholson, T. (1994). Getting off to a better start in reading and spelling: The effects of phonemic awareness instruction within a whole language program. Journal of Educational Psychology, 86, 350-359.

* Cunningham, A. (1990). Explicit versus implicit instruction in phonemic awareness. Journal of Experimental Child Psychology, 50, 429-444.

* Davidson, M.; & Jenkins, J. (1994). Effects of phonemic processes on word reading and spelling. Journal of Educational Research, 87, 148-157.

* Defior, S., & Tudela, P. (1994). Effect of phonological training on reading and writing acquisition. Reading and Writing: An Interdisciplinary Journal, 6, 299-320.

* Ehri, L., & Wilce, L. (1987b). Does learning to spell help beginners learn to read words? Reading Research Quarterly, 22, 48-65.

* Farmer, A., Nixon, M., & White, R. (1976). Sound blending and learning to read: An experimental investigation. British Journal of Educational Psychology, 46, 155-163.

* Fox, B., & Routh, D. (1976). Phonemic analysis and synthesis as word-attack skills. Journal of Educational Psychology, 68, 70-74.

* Fox, B., & Routh, D. (1984). Phonemic analysis and synthesis as word-attack skills: Revisited. Journal of Educational Psychology, 76, 1059-1064.

* Gross, J., & Garnett, J. (1994). Preventing reading difficulties: Rhyme and alliteration in the real world. Educational Psychology in Practice, 9, 235-240.

* Haddock, M. (1976). Effects of an auditory and an auditory-visual method of blending instruction on the ability of prereaders to decode synthetic words. Journal of Educational Psychology, 68, 825-831.

* Hatcher, P., Hulme, C., & Ellis, A. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. Child Development, 65, 41-57.

* Hohn, W., & Ehri, L. (1983). Do alphabet letters help prereaders acquire phonemic segmentation skill? Journal of Educational Psychology, 75, 752-762.

* Hurford, D., Johnston, M., Nepote, P., Hampton, S., Moore, S., Neal, J., Mueller, A., McGeorge, K., Huff, L., Awad, A., Tatro, C., Juliano, C., & Huffman, D. (1994). Early identification and remediation of phonological-processing deficits in first-grade children at risk for reading disabilities. Journal of Learning Disabilities, 27, 647-659.

* Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery Program. Journal of Educational Psychology, 85, 112-126.

* Kennedy, K., & Backman, J. (1993). Effectiveness of the Lindamood Auditory Discrimination in depth program with students with learning disabilities. Learning Disabilities Research and Practice, 8, 253-259.

* Korkman, M., & Peltomaa, A. (1993). Preventive treatment of dyslexia by a preschool training program for children with language impairments. Journal of Clinical Child Psychology, 22, 277-287.

* Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. Learning and Instruction, 5, 187-201.

* Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. Reading Research Quarterly, 26, 234-250.

* Lovett, M., Barron, R., Forbes, J., Cuksts, B., & Steinbach, K. (1994). Computer speech-based training of literacy skills in neurologically impaired children: A controlled evaluation. Brain and Language, 47, 117-154.

* Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. Reading Research Quarterly, 23, 263-284.

* McGuiness, D., McGuiness, C., & Donohue, J. (1995). Phonological training and the alphabet principle: Evidence for reciprocal causality. Reading Research Quarterly, 30, 830-852.

* Murray, B. (1998). Gaining alphabetic insight: Is phoneme manipulation skill or identity knowledge causal? Journal of Educational Psychology, 90, 461-475.

* O'Connor, R., & Jenkins, J. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. The Journal of Special Education, 29, 255-275.

* O'Connor, R., Jenkins, J., & Slocum, T. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. Journal of Educational Psychology, 87, 202-217.

* O'Connor, R., Notari-Syverson, A., & Vadasy, P. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. Exceptional Children, 63, 117-130.

* O'Connor, R., Notari-Syverson, A., & Vadasy, P. (1998). First-grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. Learning Disabilities Research and Practice, 13, 43-52. (followup to O'Connor et al., 1996)

* Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? Scandinavian Journal of Psychology, 24, 35-44.

* Olofsson, A., & Lundberg, I. (1985). Evaluation of long-term effects of phonemic awareness training in kindergarten: Illustrations of some methodological problems in evaluation research. Scandinavian Journal of Psychology, 26, 21-34. (Followup to Olofsson & Lundberg, 1983)

* Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergartners. Scientific Studies of Reading, 2, 301-320.

* Sanchez, E., & Rueda, M. (1991). Segmental awareness and dyslexia: Is it possible to learn to segment well and yet continue to read and write poorly? Reading and Writing: An Interdisciplinary Journal, 3, 11-18.

* Schneider, W., Kuspert, P., Roth, E., Vise, M., & Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. Journal of Experimental Child Psychology, 66, 311-340.

* Solity, J. (1996). Phonological awareness: Learning disabilities revisited? Educational & Child Psychology, 13, 103-113.

* Tangel, D., & Blachman, B. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. Journal of Reading Behavior, 24, 233-261.

* Torgesen, J., Morgan, S., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. Journal of Educational Psychology, 84, 364-370.

* Treiman, R., & Baron, J. (1983). Phonemic-analysis training helps children benefit from spelling sound rules. Memory and Cognition, 11, 382-389.

* Uhry, J., & Shepherd, M. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. Reading Research Quarterly, 28, 218-233.

* Vadasy, P., Jenkins, J., Antil, L., Wayne, S., & O'Connor, R. (1997). Community-based early reading intervention for at-risk first graders. Learning Disabilities Research and Practice, 12, 29-39.

* Vadasy, P., Jenkins, J., Antil, L., et al. (1997). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. Learning Disability Quarterly, 20, 126-139.

* Vellutino, F., & Scanlon, D. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. Merrill-Palmer Quarterly, 33, 321-363.

* Warrick, N., Rubin, H., & Rowe-Walsh, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. Annals of Dyslexia, 43, 153-173.

* Weiner, S. (1994). Effects of phonemic awareness training on low- and middle-achieving first graders' phonemic awareness and reading ability. Journal of Reading Behavior, 26, 277-300.

* Williams, J. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. Journal of Educational Psychology, 72, 1-15.

* Wilson, J., & Frederickson, N. (1995). Phonological awareness training: An evaluation. Educational & Child Psychology, 12, 68-79.

* Wise, B., King, J., & Olson, R. (1999). Training phonological awareness with and without explicit attention to articulation. Journal of Experimental Child Psychology, 72, 271-304.

* Wise, B., King, J., & Olson, R. (in press). Individual differences in benefits from computer-assisted remedial reading. Journal of Experimental Child Psychology.

Appendix B

Table 1: Dependent and Moderator Variables Included in the Meta-Analyses

OUTCOME MEASURES

1. Composite measures
    - Phonemic awareness
    - Reading
    - Spelling
2. Measures of phonemic awareness
    - Segmentation
    - Blending
    - Deletion
    - Other
3. Measures of reading
    - Standardized vs. experimenter-devised tests of word reading
    - Standardized vs. experimenter-devised tests of nonword reading
    - Reading comprehension
4. Measures of spelling
    - Standardized vs. experimenter-devised tests of spelling
5. Measure of math achievement
6. Test points
    - Immediately after training
    - First followup test (delay of 2 to 15 months)
    - Second followup test (delay of 7 to 36 months)

PROPERTIES OF PHONEMIC AWARENESS TRAINING

1. PA skills taught:
    a. Single skill; 2 skills; 3 or more skills
    b. Segmenting and blending vs. 3 or more skills
2. Use of letters: phonemes and letters manipulated vs. only phonemes manipulated
3. Training unit: individuals; small groups (2 to 7 students); classrooms
4. Identity of trainer: classroom teachers; computers; researchers/others
5. Length of training: ranged from 1 hour to 75 hours

CHARACTERISTICS OF PARTICIPANTS

1. Reader level: at-risk readers; disabled readers; normally progressing readers
2. Grade level: preschool; kindergarten; 1st grade; 2nd through 6th grades
3. Language: English; other (Dutch, Finnish, German, Hebrew, Norwegian, Spanish, Swedish)
4. Socioeconomic status: low SES; middle-to-high SES

Table 1 (continued)

FEATURES OF THE DESIGN

1. Group assignment: random; matched; non-equivalent
2. Fidelity of trainers checked vs. not checked or not reported
3. Control group: alternative treatment; no treatment
4. Size of the sample: ranged from 9 to 383 students
5. Internal validity (from Troia, 1999):
    - Percentage of criteria met
    - Number of critical flaws
6. External validity (from Troia, 1999):
    - Percentage of criteria met
    - Number of critical flaws
7. Methodological rigor (from Troia, 1999):
    - Overall ranking

CHARACTERISTICS OF THE STUDY

Year of publication (1976 to 2000)

Appendix C

Table 2: Phonemic Awareness Outcomes

Phonemic Awareness Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Time of Posttest					
Immediate	72	0.86*	No	0.79 to 0.92	ns
Followup	14	0.73*	Yes	0.61 to 0.85	
Outcome Measures of PA					
Segmentation (S)	51	0.87*	No	0.79 to 0.94	S = D > B
Blending (B)	33	0.61*	No	0.52 to 0.69	S > O
Deletion (D)	25	0.82*	No	0.73 to 0.91	B = O
Other (O)	37	0.72*	No	0.64 to 0.81	D = O
Characteristics of PA Training					
1 skill taught (1)	18	1.16*	No	0.96 to 1.36	1 = 2 > 3 +
2 skills (2)	24	1.03*	No	0.92 to 1.14	
3 or more skills (3)	30	0.70*	No	0.61 to 0.78	
Blend & segment only	18	0.81*	No	0.67 to 0.95	ns
3 or more skills	30	0.70*	No	0.61 to 0.78	
Letters manipulated	39	0.89*	No	0.80 to 0.98	ns
Letters not manipulated.	33	0.82*	No	0.73 to 0.91	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Individual child (I)	24	0.60*	Yes	0.47 to 0.72	S > I = C
Small groups (S)	35	1.38*	No	1.26 to 1.50	
Classrooms (C)	13	0.67*	No	0.57 to 0.76	
Length of training					
1 to 4.5 hrs (1)	15	0.61*	Yes	0.41 to 0.81	2 = 3 > 1 = 4
5 to 9.3 hrs (2)	24	1.37*	No	1.23 to 1.51	
10 to 18 hrs (3)	9	1.14*	No	0.97 to 1.32	
20 to 75 hrs (4)	22	0.65*	No	0.56 to 0.74	
Characteristics of Trainers					
Classroom teachers (CT)	19	0.78*	No	0.70 to 0.87	RO > CT
Researchers & others (RO)	53	0.94*	No	0.84 to 1.03	
Computers (Com)	8	0.66*	Yes	0.52 to 0.85	O > Com
Others (O)	64	0.89*	No	0.82 to 0.96	
Characteristics of Participants					
Reading level					
At risk (A)	15	0.95*	No	0.76 to 1.14	A = N > D
Disabled (D)	15	0.62*	No	0.48 to 0.75	
Normal progress (N)	42	0.93*	No	0.85 to 1.01	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Grade					
Preschool (Pre)	2	2.37*	No	1.93 to 2.81	Pre > K > 1 = 2
Kindergarten (K)	39	0.95*	No	0.87 to 1.04	
1st (1)	15	0.48*	Yes	0.31 to 0.64	
2nd-6th (2)	16	0.70*	Yes	0.56 to 0.83	
Socioeconomic status					
Low	12	1.07*	No	0.93 to 1.20	ns
Mid & High	17	1.02*	No	0.87 to 1.18	
Language					
English (E)	61	0.99*	No	0.90 to 1.07	E > O
Other (O)	11	0.65*	Yes	0.55 to 0.76	
Characteristics of Design					
Random assignment	33	0.87*	No	0.77 to 0.97	ns
Matched	18	0.92*	No	0.75 to 1.09	
Non-equivalent	21	0.83*	No	0.73 to 0.92	
Fidelity checked (FCh)	29	0.66*	No	0.56 to 0.75	Not > FCh
Not checked (Not)	43	1.02*	No	0.93 to 1.11	
Treated controls	38	0.89*	No	0.79 to 0.99	ns
Untreated controls	34	0.83*	No	0.75 to 0.92	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Size of sample					
9 to 22 students (1)	15	1.37*	No	1.09 to 1.66	1 = 3 > 2 = 4
24 to 30 students (2)	22	0.70*	No	0.53 to 0.87	
31 to 53 students (3)	13	1.10*	No	0.90 to 1.30	
56 to 383 students (4)	22	0.82*	No	0.74 to 0.89	
Characteristics of Study					
Year of publication					
1976-1985 (1)	10	0.73*	Yes	0.53 to 0.94	3 > 1 = 2 = 4
1986-1990 (2)	16	0.72*	No	0.59 to 0.85	
1991-1995 (3)	31	1.18*	No	1.07 to 1.30	
1996-2000 (4)	15	0.70*	No	0.59 to 0.81	

* indicates that effect size was significantly greater than zero at $p < 0.05$. ns indicates not significantly different from zero.

Table 3: Reading Outcomes

Reading Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Outcome Measures					
Time of posttest					
Immediate (Im)	90	0.53*	No	0.47 to 0.58	Im = 1 > 2
1st followup (1)	35	0.45*	No	0.36 to 0.54	
2nd followup (2)	8	0.23*	No	0.11 to 0.34	
Type of word test					
Experimenter (E)	58	0.61*	No	0.54 to 0.69	E > S
Standardized (S)	39	0.33*	No	0.24 to 0.42	
Type of pseudoword test					
Experimenter	47	0.56*	No	0.48 to 0.64	ns
Standardized	8	0.49*	Yes	0.29 to 0.69	
Reading comprehension	18	0.32*	No	0.18 to 0.46	
Math achievement	15	0.03ns	No	-0.11 to 0.16	
Characteristics of PA Training					
Immediate posttest					
1 skill taught (1)	32	0.71*	No	0.58 to 0.84	1 = 2 > 3 +
2 skills taught (2)	29	0.79*	No	0.69 to 0.89	
3 or more skills (3)	29	0.27*	Yes	0.19 to 0.35	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Followup posttest					
1 skill taught (1)	11	0.55*	Yes	0.37 to 0.73	2 > 1 > 3 +
2 skills (2)	9	1.28*	No	0.56 to 0.89	
3 or more skills (3)	15	0.23*	Yes	0.11 to 0.37	
Blend & segment only (BS)	19	0.67*	No	0.54 to 0.81	BS > 3 +
3 or more skills (3)	29	0.27*	Yes	0.19 to 0.35	
Immediate posttest					
Letters manipulated (L)	48	0.67*	No	0.59 to 0.75	L > NoL
Letters not manipulated (NoL)	42	0.38*	No	0.30 to 0.46	
Followup posttest					
Letters manipulated (L)	16	0.59*	No	0.45 to 0.74	L > NoL
Letters not manipulated (NoL)	19	0.36*	No	0.25 to 0.47	
Immediate posttest					
Individual child (I)	32	0.45*	Yes	0.34 to 0.57	S > I = C
Small groups (S)	42	0.81*	No	0.71 to 0.92	
Classrooms (C)	16	0.35*	No	0.26 to 0.44	
Followup posttest					
Individual child (I)	7	0.33*	Yes	0.11 to 0.55	S > I = C

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Small groups (S)	18	0.83*	No	0.66 to 1.00	
Classrooms (C)	10	0.30*	Yes	0.18 to 0.42	
Length of training					
1 to 4.5 hrs (1)	17	0.61*	Yes	0.42 to 0.79	1 = 2 = 3 > 4
5 to 9.3 hrs (2)	23	0.76*	No	0.62 to 0.89	
10 to 18 hrs (3)	19	0.86*	No	0.72 to 1.00	
20 to 75 hrs (4)	25	0.31*	No	0.22 to 0.39	
Characteristics of Trainers					
Immediate posttest					
Classroom teachers (CT)	22	0.41*	No	0.33 to 0.49	RO > CT
Researchers & others (RO)	68	0.64*	No	0.56 to 0.73	
Followup posttest					
Classroom teachers (CT)	12	0.32*	Yes	0.20 to 0.43	RO > CT
Researchers & others (RO)	23	0.63*	No	0.49 to 0.77	
Computers (Com)	8	0.33*	Yes	0.16 to 0.49	O > Com
Others (O)	82	0.55*	No	0.49 to 0.61	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Participants					
Reading level: Immediate posttest					
At risk (A)	27	0.86*	No	0.72 to 1.00	A > D = N
Disabled (D)	17	0.45*	Yes	0.32 to 0.57	
Normal progress (N)	46	0.47*	No	0.39 to 0.54	
Reading level: Followup posttest					
At risk	15	1.33*	No	1.10 to 1.56	A > D = N
Disabled	8	0.28*	Yes	0.10 to 0.46	
Normal progress	12	0.30*	Yes	0.19 to 0.42	
Grade					
Preschool (Pre)	7	1.25*	No	1.01 to 1.50	Pre > K = 1 = 2
Kindergarten (K)	40	0.48*	No	0.40 to 0.56	
1st (1)	25	0.49*	Yes	0.36 to 0.62	
2nd-6th (2)	18	0.49*	Yes	0.35 to 0.62	
Socioeconomic status					
Low (L)	11	0.45*	No	0.33 to 0.58	MH > L
Mid & High (MH)	29	0.84*	No	0.72 to 0.96	
Language					
Immediate posttest					
English (E)	72	0.63*	No	0.55 to 0.70	E > O
Other (O)	18	0.36*	No	0.27 to 0.46	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Followup posttest					
English (E)	17	0.42*	Yes	0.28 to 0.56	ns
Other (O)	18	0.47*	No	0.35 to 0.59	
Characteristics of Design					
Random assignment (R)	46	0.63*	No	0.54 to 0.72	R > N
Matched (M)	22	0.57*	Yes	0.43 to 0.72	M = all
Nonequivalent (N)	20	0.40*	No	0.31 to 0.49	
Fidelity checked (FCh)	31	0.43*	No	0.34 to 0.53	Not > FCh
Not checked (Not)	59	0.59*	No	0.51 to 0.66	
Immediate posttest					
Treated controls (T)	54	0.65*	No	0.56 to 0.73	T > U
Untreated controls (U)	36	0.41*	No	0.33 to 0.49	
Followup posttest					
Treated contols (T)	20	0.62*	No	0.48 to 0.75	T > U
Untreated controls (U)	15	0.32*	Yes	0.20 to 0.44	
Size of sample					
9 to 22 students (1)	24	0.72*	No	0.51 to 0.92	1 = 3 > 4
24 to 30 students (2)	22	0.54*	Yes	0.37 to 0.70	2 = 1, 4
31 to 53 students (3)	22	0.91*	No	0.76 to 1.05	3 > 2

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
56 to 383 students (4)	22	0.40*	No	0.33 to 0.48	
Characteristics of Study					
Year of publication					
1976-1985 (1)	20	0.77*	No	0.62 to 0.93	1 = 3 > 2 = 4
1986-1990 (2)	16	0.36*	Yes	0.24 to 0.49	
1991-1995 (3)	41	0.77*	No	0.67 to 0.87	
1996-2000 (4)	13	0.21*	Yes	0.11 to 0.32	

* indicates that effect size was significantly greater than zero at $p < 0.05$.

ns indicates not significantly different from zero.

Table 4: Spelling Outcomes

Spelling Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Outcome Measures					
Time of Posttest					
Immediate (Im)	39	0.59*	No	0.49 to 0.68	Im > 1 = 2
1st followup (1)	17	0.37*	Yes	0.26 to 0.48	
2nd followup (2)	6	0.20*	No	0.08 to 0.32	
Type of spelling test					
Experimenter (E)	24	0.75*	No	0.62 to 0.89	E > S
Standardized (S)	20	0.41*	No	0.29 to 0.53	
Characteristics of PA Training					
1 skill taught (1)	17	0.74*	No	0.56 to 0.92	1 = 2 > 3 +
2 skills (2)	12	0.87*	Yes	0.71 to 1.03	
3 or more skills (3)	10	0.23*	No	0.07 to 0.38	
Blend & segment only (BS)	7	0.79*	Yes	0.49 to 1.09	BS > 3 +
3 or more skills (3)	10	0.23*	No	0.07 to 0.38	
Letters manipulated (L)	27	0.61*	No	0.50 to 0.72	L > NoL
Letters not used (NoL)	12	0.34*	No	0.25 to 0.42	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Individual child (I)	14	0.36*	No	0.20 to 0.52	S > I
Small groups (S)	20	0.77*	No	0.63 to 0.90	C = all
Classrooms (C)	5	0.56*	No	0.33 to 0.78	
Length of training					
1 to 4.5 hrs (1)	0	Ñ	Ñ	Ñ	
5 to 9.3 hrs (2)	8	1.13*	Yes	0.86 to 1.39	2 = 3 > 4
10 to 18 hrs (3)	10	0.87*	No	0.69 to 1.05	
20 to 75 hrs (4)	18	0.32*	No	0.19 to 0.45	
Characteristics of Trainers					
Classroom teachers (CT)	9	0.74*	No	0.58 to 0.90	CT > RO
Researchers & others (RO)	30	0.51*	No	0.39 to 0.62	
Computers (Com)	6	0.09ns	Yes -0.10 to 0.28	O > Com	
Others (O)	33	0.74*	No	0.63 to 0.85	
Characteristics of Participants					
Reading level					
At risk (A)	13	0.76*	No	0.54 to 0.98	A = N > D
Disabled (D)	11	0.15ns	Yes	-0.00 to 0.31	
Normal progress (N)	15	0.88*	No	0.74 to 1.02	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Grade					
Preschool (P)	0	Ñ	Ñ	Ñ	
Kindergarten (K)	15	0.97*	No	0.82 to 1.13	K > 1 > 2
1st (1)	16	0.52*	No	0.37 to 0.68	
2nd-6th (2)	8	0.14ns	Yes	-0.04 to 0.33	
Socioeconomic status					
Low (L)	6	0.76*	Yes	0.57 to 0.95	MH > L
Mid & High (MH)	9	1.17*	No	0.88 to 1.47	
Language					
English	32	0.60*	No	0.49 to 0.70	ns
Other	7	0.55*	Yes	0.31 to 0.78	
Characteristics of Design					
Random assignment (R)	17	0.37*	No	0.23 to 0.50	M = N > R
Matched (M)	12	0.73*	No	0.52 to 0.93	
Nonequivalent (N)	10	0.86*	Yes	0.69 to 1.04	
Fidelity checked (FCh)	15	0.44*	No	0.30 to 0.59	Not > FCh
Not checked (Not)	24	0.69*	No	0.57 to 0.81	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Treated controls (T)	24	0.43*	No	0.30 to 0.55	U > T
Untreated controls (U)	15	0.82*	No	0.67 to 0.96	
Size of sample					
9 to 22 students (1)	15	0.85*	Yes	0.59 to 1.10	2 > all
24 to 30 students (2)	3	1.68*	Yes	1.15 to 2.21	1 > 4
31 to 53 students (3)	8	0.75*	No	0.51 to 0.98	3 = 1, 4
56 to 383 students (4)	13	0.45*	No	0.34 to 0.56	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero

Appendix D

Table 5: Results

Mean Effect Sizes (d) With Reading Disabled Comparisons Removed from the Data Base and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
SPELLING OUTCOMES					
PA Skills Taught					
1 skill taught	14	0.77*	No	0.58 to 0.96	ns
2 skills taught	11	0.89*	Yes	0.72 to 1.05	
3 or more skills	3	0.93*	No	0.52 to 1.33	
Blend & segment only	6	0.85*	Yes	0.54 to 1.16	ns
3 or more skills	3	0.93*	No	0.52 to 1.33	
Letters manipulated (L)	17	1.00*	Yes	0.85 to 1.15	L > NoL
Letters not manipulated (NoL)	11	0.57*	No	0.37 to 0.76	
Training Unit					
Individual child (I)	8	1.00*	No	0.71 to 1.28	I = S > C
Small groups (S)	15	0.94*	Yes	0.78 to 1.10	
Classrooms (C)	5	0.56*	No	0.33 to 0.78	
Length of training					
1 to 4.5 hrs	0	0	Ñ	Ñ	
5 to 9.3 hrs	8	1.13*	Yes	0.86 to 1.39	ns
10 to 18 hrs	8	0.91*	No	0.73 to 1.10	
20 to 75 hrs	9	0.75*	Yes	0.50 to 1.01	

Table 5: Results (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Trainer					
Classroom teachers	8	0.74*	No	0.58 to 0.91	ns
Researchers and others	20	0.96*	No	0.79 to 1.14	
Grade					
Preschool (Pre)	0				
Kindergarten (K)	15	0.97*	No	0.82 to 1.13	K > 1
1st (1)	13	0.66*	No	0.48 to 0.85	
2nd-6th (2)	0				
Language					
English (E)	22	0.95*	No	0.82 to 1.09	E > O
Other (O)	6	0.51*	Yes	0.28 to 0.75	
PHONEMIC AWARENESS OUTCOMES					
Letters manipulated (L)	25	1.11*	No	0.99 to 1.23	L > NoL
Letter not manipulated (NoL)	32	0.83*	No	0.73 to 0.92	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Appendix E

Table 6

Phonemic Awareness Outcomes: Mean Effect Sizes (d) Asssociated With Troia's Indicators of Methodological Rigor and Tests to Determine Whether Effect Sizes Were Significantly Greater than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Internal Validity				
% of criteria met				
24-40% (1)	10	0.67*	Yes	2 = 4 > 1
47% (2)	5	1.35*	No	4 > 3
53% (3)	14	0.95*	No	2 = 3
59-82% (4)	14	1.66*	No	
Critical Flaws				
1-2 (1)	18	1.63*	No	1 > 3 > 2
3 (2)	14	0.57*	Yes	
4-5 (3)	11	0.97*	No	
External Validity				
% of criteria met				
47-53% (1)	10	0.92*	No	4 > 1 = 2
56-60% (2)	14	0.81*	No	3 = 2, 4, 1
63-67% (3)	8	1.13*	No	
73-81% (4)	11	1.40*	No	
Critical flaws				
0 flaws	13	1.69*	No	0 > all
1	8	0.96*	No	1 = 2 = 3
2	13	0.61*	Yes	
3	9	0.97*	No	

Table 6 (continued)

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Ranking				
High rigor (1-12) (1)	15	1.56*	No	1 = 2 > 3
Mid (13-24) (2)	11	1.40*	No	
Low (25-36) (3)	17	0.69*	Yes	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Table 7

Reading Outcomes: Mean Effect Sizes (d) Asssociated With Troia's Indicators of Methodological Rigor and Tests to Determine Whether Effect Sizes Were Significantly Greater than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Internal Validity				
% of criteria met				
24-40% (1)	11	0.49*	No	2 > 1
47% (2)	15	0.85*	No	4 > 1
53% (3)	16	0.63*	No	2 = 3 = 4
59-82% (4)	14	0.83*	No	1 = 3
Critical Flaws				
1-2 (1)	22	0.99*	No	1 > 2 = 3
3 (2)	18	0.59*	Yes	
4-5 (3)	16	0.56*	No	
External Validity				
% of criteria met				
47-53% (1)	16	0.98*	No	1 > 2, 3
56-60% (2)	14	0.58*	Yes	1 = 4
63-67% (3)	15	0.61*	No	2 = 3 = 4
73-81% (4)	11	0.66*	No	

Table 7 (continued)

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Critical Flaws				
0 flaws	17	0.90*	No	0 = 3 > 1
1	11	0.51*	No	2 = all
2	17	0.57*	Yes	
3	11	0.92*	No	
Ranking				
High rigor (1-12) (1)	19	1.00*	No	1 > 2 = 3
Mid (13-24) (2)	14	0.61*	Yes	
Low (25-36) (3)	23	0.58*	No	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Appendix F

Studies in the Phonemic Awareness (PA) Database, Their Characteristics, and Effect Sizes

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Ball & Blachman, 1991													89			
01 - Segment & categ. + let vs. Language, LS	2	Yes	SmG	Other	9.33	Nor	K	Engl		R	Yes	59		1.49	0.71	0.87
02 - Segment & categ. + let vs. No treatment	2	Yes	SmG	Other	9.33	Nor	K	Engl		R	Yes	59		1.64	0.98	0.83
Barker & Torgesen, 1995													36			
03 - Mult. PA on computers vs. math on computers	3+	No	Ind	Comp	13.33	AR	1st	Engl		R	No	36		0.48	0.22	
Bentin & Leshem, 1993													91			
04 - Segment & categ. vs. Language	2	No	SmG	Other	10	AR	K	Hebr	M-H	R	No	50			4.21	
05 - Segment & categ. vs. No treatment	2	No	SmG	Other	10	AR	K	Hebr	M-H	R	No	41			4.33	
06 - Segment & categ. + let vs. Language	2	Yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	50			2.1	
07 - Segment & categ. + let vs. No treat	2	Yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	41			2.17	
Blachman et al., 1994													159			
08 - Segment & categ. + let vs. No treat.	2	Yes	SmG	Teach	12.3	Nor	K	Engl	Lo	NE	No	159		1.83	0.65	0.94
Bradley & Bryant, 1983, 1985													65			
09 - Phon. categ. vs. Semantic categ.	1	No	Ind	Other	11.67	AR	1st	Engl		M/R	No	39			0.5	0.39
10 - Phon. categ. vs. No treatment	1	No	Ind	Other	11.67	AR	1st	Engl		M/R	No	26			0.86	1
11 - Phon. categ. + let vs. Semantic categ.	1	Yes	Ind	Other	11.67	AR	1st	Engl		M/R	No	39			1.17	1.59
12 - Phon. categ. + let vs. No treatment	1	Yes	Ind	Other	11.67	AR	1st	Engl		M/R	No	26			1.53	2.18

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Brady et al., 1994	.	.	.	.	.	.	.	.	.	.	.	.	42	.	.	.
13 - Mult. PA vs. No treatment	3+	No	Clas	Teach	18	AR	K	Engl	Lo	NE	Yes	42		0.46	0.47	0.23
Brennan & Ireson, 1997	.	.	.	.	.	.	.	.	.	.	.	.	24	.	.	.
14 - Segment & blend vs. No treatment	2	No	Clas	Teach	48	Nor	K	Engl	M-H	NE	Yes	24		3.92	1.17	2.17
Bus, 1986	.	.	.	.	.	.	.	.	.	.	.	.	201	.	.	.
15 - Segment & blend, LS vs. Pre-read prep., LS	2	No	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	130	.	0.55	0.54	.
16 - Segment & blend + let vs. Pre-read prep., LS	2	Yes	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	134	.	0.25	0.35	.
Byrne & Fielding-Barnsley, 1991, '93, '95	.	.	.	.	.	.	.	.	.	.	.	.	126	.	.	.
17 - Phon. categ. + let vs. Semantic categ.	1	Yes	SmG	Other	6	Nor	Pre	Engl	M-H	R	No	126		3.14	1.61	.34*
Castle, et al., 1994, Experiment 2	.	.	.	.	.	.	.	.	.	.	.	.	51	.	.	.
18 - Mult. PA + let vs. Language	3+	Yes	SmG	Other	5	Nor	K	Engl	M-H	M/R	No	34		3.81	1.06	1.27
19 - Mult. PA + let vs. No treatment	3+	Yes	SmG	Other	5	Nor	K	Engl	M-H	M/R	No	34		2.62	1.09	1.73
Cunningham, 1990	.	.	.	.	.	.	.	.	.	.	.	.	84	.	.	.
20 - Segment & blend vs. Stories	2	No	SmG	Other	6	Nor	K	Engl	M-H	M/R	No	28	.	1.62	0.42	.
21 - Segment, blend, meta. vs. Stories	2	No	SmG	Other	6	Nor	K	Engl	M-H	M/R	No	28	.	2.3	0.56	.
22 - Segment & blend vs. Stories	2	No	SmG	Other	6	Nor	1st	Engl	M-H	M/R	No	28	.	0.99	0.08	.
23 - Segment, blend, meta. vs. Stories	2	No	SmG	Other	6	Nor	1st	Engl	M-H	M/R	No	28	.	1.27	0.51	.
Davidson & Jenkins, 1994	.	.	.	.	.	.	.	.	.	.	.	.	40	.	.	.
24 - Segment, LS vs. Stories, LS	1	No	SmG	Other	8.33	Nor	K	Engl		NE	No	20	.	8	1.58	1.6

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
25 - Blend, LS vs. Stories, LS	1	No	SmG	Other	8.33	Nor	K	Engl		NE	No	20		3.11	0.71	0.49
26 - Segment & blend, LS vs. Stories, LS	2	No	SmG	Other	8.33	Nor	K	Engl		NE	No	20		3.93	1.56	1.13
Defior & Tudela, 1994													43			
27 - Categ. + let vs. Semantic categ.	1	Yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22			0.82	1.44
28 - Categ. + let vs. Hand manipulation	1	Yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22			0.73	1.03
29 - Categ. vs. Semantic categ.	1	No	SmG	Other	30	AR	1st	Span	M-H	R	No	21			0.18	0.36
30 - Categ. vs. Hand manipulation	1	No	SmG	Other	30	AR	1st	Span	M-H	R	No	21			0.14	0.02
Ehri & Wilce, 1987													20			
31 - Segment + let vs. LS	1	Yes	Ind	Other	5.6	Nor	K	Engl	M-H	M/R	No	20		1.99	0.97	2.59
Farmer et al., 1976													60			
32 - Blend + let vs. Label pictures	1	Yes	Ind	Other		Nor	1st	Engl		R	No	20		0.78	0.96	
33 - Blend + let vs. Label pictures	1	Yes	Ind	Other		Nor	K	Engl		R	No	40		0.63	0.35	
Fox & Routh, 1976													40			
34 - Read training with blend vs. Without blend	1	No	Ind	Other	1	Nor	Pre	Engl	M-H	R	No	20			1.61	
35 - Read training with blend vs. Without blend	1	No	Ind	Other	1	AR	Pre	Engl	M-H	R	No	20			-0.1	
Fox & Routh, 1984													31			
36 - Onset-rime, LS vs. No treat., LS	1	No	SmG	Other	5	AR	K	Engl		R	No	21		0.75	-0.19	
37 - Onset-rime, LS vs. No treat., LS	1	No	SmG	Other	5	AR	K	Engl		R	No	21		1.6	3.6	

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Gross & Garnet, 1994													12			
38 - Categ. vs. No treatment	1	No	SmG	Other		AR	K	Engl	Lo	M/R	No	12			2.29*	.60*
Haddock, 1976													80			
39 - Blend, LS vs. LS	1	No	Clas	Teach	2.5	Nor	Pre	Engl		NE	No	53			0.92	
40 - Blend + let vs. LS	1	Yes	Clas	Teach	2.5	Nor	Pre	Engl		NE	No	48			1.67	
Hatcher et al., 1994													124			
41 - Mult. PA vs. No treatment	3+	No	Ind	Other	20	RD	1st	Engl		M/R	Yes	61		0.64	0.13	0.25
42 - Mult. PA + let in Read Rec. vs. Read Rec.	3+	Yes	Ind	Other	20	RD	1st	Engl		M/R	Yes	63		0.24	0.31	0.31
Hohn & Ehri, 1983													24			
43 - Segment vs. No treatment	1	No	Ind	Other	2.58	Nor	K	Engl		M/R	No	16		0.77	0.2	
44 - Segment + let vs. No treatment	1	Yes	Ind	Other	2.58	Nor	K	Engl		M/R	No	16		1.3	0.68	
Hurford et al., 1994													99			
45 - Blend & deletion + let vs. No treat.	2	Yes	Ind	Comp	12	AR	1st	Engl	M-H	M/R	No	99		0.61	0.49	
Iversen & Tunmer, 1993													64			
46 - Mult. PA + let in Read Rec. vs. Read Rec.	3+	Yes	Ind	Other	20.88	RD	1st	Engl		M/R	Yes	64		-0.33	0.42	-0.02
Kennedy & Backman, 1993													20			
47 - Mult. PA + let vs. No treatment	3+	Yes	SmG	Teach	75	RD	2nd+	Engl		M/R	Yes	20		1.43	0.39	0.53
Korkman & Peltoma, 1993													46			
48 - Blend & categ. + let vs. Speech therapy	2	Yes	SmG	Other		AR	K	Fin		NE	No	46			.60*	.67*

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Kozminsky & Kozminsky, 1995													61			
49 - Mult. PA vs. Visual motor integration	3+	No	Clas	Teach	21.33	Nor	K	Hebr	Lo	NE	No	61		0.24	.57*	
Lie, 1991													208			
50 - Categ. vs. Conceptual	1	No	Clas	Teach		Nor	1st	Norw		R	No	96			0.21	0.22
51 - Segment vs. Conceptual	1	No	Clas	Teach		Nor	1st	Norw		R	No	102			0.62	0.67
Lovett et al., 1994													19			
52 - Segment & blend + let vs. Whole word	2	Yes	Ind	Comp	18	RD	2nd+	Engl		R	No	13			1	0.02
53 - Onset-rime + let vs. Whole word	1	Yes	Ind	Comp	18	RD	2nd+	Engl		R	No	13			0.53	0.15
Lundberg et al., 1988													383			
54 - Mult. PA vs. No treatment	3+	No	Clas	Teach	48	Nor	K	Dan	Lo	NE	No	383		0.74	0.19	.60*
McGuinness et al., 1995, Study 2													42			
55 - Mult. PA + let in Montessori vs. No treat.	3+	Yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27		0.15	1.11	
56 - Mult. PA + let in whole lang. vs. No treat.	3+	Yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27		0.37	1.22	
Murray, 1998													48			
57 - Categ., LS vs. Language, LS	1	No	SmG	Other	4.5	Nor	K	Engl		R	Yes	30		-0.11	0.27	
58 - Segment & blend, LS vs. Language, LS	2	No	SmG	Other	4.5	Nor	K	Engl		R	Yes	30		0.41	0.07	
O'Connor & Jenkins, 1995													10			
59 - Segment + let to spell vs. LS, read	1	Yes	Ind	Other	3.33	AR	K	Engl		M/R	No	10		0.41	0.9	1.24
O'Connor et al., 1995													67			
60 - Segment & blend, LS vs. LS	2	No	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45		2.69	1.64	

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
61 - Mult. PA, LS vs. LS	3+	No	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45		2.42	0.52	
O'Connor et al., 1996, 1998													80			
62 - Segment & blend + let vs. No treat.	2	Yes	Clas	Teach	20	Nor	K	Engl		NE	Yes	66		0.62	0.11	0.73
63 - Segment & blend + let vs. No treat.	2	Yes	SmG	Teach	20	AR	K	Engl		NE	Yes	14		0.03	0.99	0.97
Olofsson & Lundberg, 1983, 1985													48			
64 - Mult. PA scheduled vs. Nonverbal tasks	3+	No	Clas	Teach	12.25	Nor	K	Swed		NE	Yes	38		0.7	0.28	-.07*
65 - Mult. PA scheduled vs. No treatment	3+	No	Clas	Teach	12.25	Nor	K	Swed		NE	Yes	26		0.27	-0.37	0.16*
Reitsma & Wesseling, 1998													70			
66 - Blend on computers vs. Vocab. comput.	1	No	Ind	Comp	4	Nor	K	Dutch		NE	No	25		0.23	.42*	-.11*
67 - Blend on computers vs. No treatment	1	No	Ind	Comp	4	Nor	K	Dutch		R	No	56		0.74	.27*	.28*
Sanchez & Rueda, 1991													9			
68 - Segment + let vs. Percept-motor	1	Yes	SmG	Other	40	RD	2nd+	Span		R	No	9		2.19	-0.05	2.09
Schneider et al., 1997													702			
69 - Mult. PA vs. No treatment	3+	No	Clas	Teach	43.75	Nor	K	Germ		NE	No	371		0.7	0.22	.27*
70 - Mult. PA vs. No treatment	3+	No	Clas	Teach	20	Nor	K	Germ		NE	Yes	331		0.82	0.05	.38*
Solity, 1996													24			
71 - Segment & blend vs. Story	2	No	SmG	Other	14.75	Nor	Pre	Engl		M/R	Yes	24		0.52	1.18	
Tangel & Blachman, 1992													149			
72 - Segment & categ. + let vs. No treat.	2	Yes	SmG	Teach	13.2	Nor	K	Engl	Lo	NE	No	149		1.81	0.67	0.94

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Torgesen et al., 1992													48			
73 - Segment & blend. LS vs. Story. LS	2	No	SmG	Other	7	AR	K	Engl	Lo	M/R	No	31		1.87	1.22	
74 - Blend. LS vs. Story. LS	1	No	SmG	Other	7	AR	K	Engl	Lo	M/R	No	32		1.82	-0.05	
Treiman & Baron, 1983													28			
75 - Onset-rime vs. Repeat syllables	1	No	Ind	Other		Nor	Pre	Engl	M-H		Yes	8			0.62	
76 - Onset-rime vs. Repeat syllables	1	No	Ind	Other		Nor	K	Engl	M-H		Yes	20			0.13	
Uhry & Shepherd, 1993													22			
77 - Seg. & blend + let vs. Text	2	Yes	SmG	Comp	17.33	Nor	1st	Engl	M-H	R	No	22		1.45	1.07	0.77
Vadasy et al., 1997 (LDRP)													35			
78 - Segment & categ. + let vs. No treat.	2	Yes	Ind	Other	54	AR	1st	Engl	Lo	R	Yes	35		0.74	0.44	0.67
Vadasy et al., 1997 (LDQ)													40			
79 - Segment & blend + let vs. No treat.	2	Yes	Ind	Other	50	AR	1st	Engl	Lo	R	Yes	40		0.42	0.27	0.4
Vellutino & Scanlon, 1987, Experiment 2													240			
80 - Mult. PA + let vs. No treatment	3+	Yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		1.15	0.72	
81 - Mult. PA, let, word vs. Word	3+	Yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		0.74	0.3	
82 - Mult. PA + let vs. No treatment	3+	Yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		0.33	0.47	
83 - Mult. PA, let, word vs. Word	3+	Yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		1.1	0.71	
84 - Mult. PA + let vs. No treatment	3+	Yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		0.89	0.49	
85 - Mult. PA, let, word vs. Word	3+	Yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		1.01	0.48	
86 - Mult. PA + let vs. No treatment	3+	Yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		-0.07	0.33	
87 - Mult. PA, let, word vs. Word	3+	Yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		0.66	0.52	

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Warrick et al., 1993, Study II													28			
88 - Segment vs. No treatment	1	No	SmG	Other	5.33	AR	K	Engl		NE	No	28		0.67	1.30*	81*
Weiner, 1994													36			
89 - Mult. PA vs. No treatment	3+	No	SmG	Other	5	AR	1st	Engl	M-H	R	No	10		0.81	0.17	
90 - Mult. PA vs. No treatment	3+	No	SmG	Other	5	Nor	1st	Engl	M-H	R	No	26		0.17	-0.06	
Williams, 1980													204			
91 - Segment & blend + let vs. No treat.	2	Yes	Clas	Teach	62.83	RD	2nd+	Engl		NE	Yes	102		0.35	1.05	
92 - Segment & blend + let vs. No treat.	2	Yes	Clas	Teach	28.13	RD	2nd+	Engl		R	Yes	102		1.11	0.97	
Wilson & Frederickson, 1995													48			
93 - Onset-rime + let vs. No treatment	1	Yes	SmG	Other	26.67	RD	2nd+	Engl	Lo	NE	Yes	48		0.12	0.47	0.49
Wise et al., 1999													122			
94 - Mult. PA w/ artic. + let vs. Artic., LS	3+	Yes	SmG	Comp	42	RD	2nd+	Engl		R	Yes	80		0.65	0.15	0.05
95 - Mult. PA + let vs. Artic., LS	3+	Yes	SmG	Comp	42	RD	2nd+	Engl		R	Yes	85		0.66	0.28	0.3
Wise et al., in press													200			
96 - Mult. PA + let vs. Recip. Teach	3+	Yes	Ind	Comp	24.98	RD	2nd+	Engl		R	No	200		0.77	0.23	-0.05
Abreviations:																

LS = Letter-sound training provided separately

Mult. = Multiple PA in 3 or more skills

Categ. = Categorization or identity training

Recip. Teach = Reciprocal teaching strategies learned and applied to reading

Meta = Metacognitive activities to understand purposes, use of PA

* = Effect sizes were drawn from followup test points

Read Rec. = Reading Recovery* program

CHAPTER
2

Alphabetics

Part II

Phonics Instruction

PART II: PHONICS INSTRUCTION

Executive Summary

Introduction

Learning to read is a complex task for beginners. They must coordinate many cognitive processes to read accurately and fluently, including recognizing words, constructing the meanings of sentences and text, and retaining the information read in memory. An essential part of the process for beginners involves learning the alphabetic system, that is, letter-sound correspondences and spelling patterns, and learning how to apply this knowledge in their reading. Systematic phonics instruction is a way of teaching reading that stresses the acquisition of letter-sound correspondences and their use to read and spell words (Harris & Hodges, 1995). Phonics instruction is designed for beginners in the primary grades and for children having difficulty learning to read.

In teaching phonics explicitly and systematically, several different instructional approaches have been used. These include synthetic phonics, analytic phonics, embedded phonics, analogy phonics, onset-rime phonics, and phonics through spelling. Although all explicit, systematic phonics approaches use a planned, sequential introduction of a set of phonic elements along with teaching and practice of those elements, they differ across a number of other features. For example, the content covered ranges from a limited to an elaborate set of letter-sound correspondences and phonics generalizations. In addition, the application procedures taught to children vary. Synthetic phonics programs teach children to convert letters into sounds or phonemes and then blend the sounds to form recognizable words. Analytic phonics avoids having children pronounce sounds in isolation to figure out words. Rather children are taught to analyze letter-sound relations once the word is identified. Phonics-through-spelling programs teach children to transform sounds into letters to write words. Phonics in context approaches teach children to use sound-letter correspondences along with context cues to identify unfamiliar words they encounter in text. Analogy phonics programs teach children to use parts of written words they already know to identify new words. The distinctions between systematic phonics approaches are not absolute, however, and some phonics programs combine two or more of these types of instruction. In addition, these approaches differ with respect to the extent that controlled vocabulary (decodable text) is used for practicing reading connected text. Although differences exist, the hallmark of systematic phonics programs is that they delineate a planned, sequential set of phonic elements and they teach these elements explicitly and systematically. The goal in all phonics programs is to enable learners to acquire sufficient knowledge and use of the alphabetic code so that they can make normal progress in learning to read and comprehend written language.

The purpose of this report is to examine the research evidence concerning systematic phonics instruction. The research literature was searched to identify experiments that compared the reading performance of children who had received systematic phonics instruction to the performance of children given nonsystematic phonics or no phonics instruction. The National Reading Panel (NRP) sought answers to the following questions:

- Does systematic phonics instruction help children learn to read more effectively than nonsystematic phonics instruction or instruction teaching no phonics?
- Are some types of phonics instruction more effective than others? Are some specific phonics programs more effective than others?
- Is phonics instruction more effective when students are taught individually, in small groups, or as whole classes?
- Is phonics instruction more effective when it is introduced in kindergarten or 1st grade to students not yet reading or in later grades after students have begun to read?
- Is phonics instruction beneficial for children who are having difficulty learning to read? Is it effective in preventing reading failure among children who are at risk for developing reading problems in the future? Is it effective in remediating reading

difficulties among children who have not made normal progress in learning to read?

- Does phonics instruction improve children's ability to read and comprehend text as well as their decoding and word-reading skills?
- Does phonics instruction have an impact on children's growth in spelling?
- Is phonics instruction effective with children at different socioeconomic (SES) levels?
- Does the type of instruction given to control groups as part of a study to evaluate phonics make a difference?
- If phonics instruction is found to be more effective than less-phonics or no-phonics instruction, were the experiments that showed these effects well designed or poorly designed?

Beginning reading programs that do not teach phonics explicitly and systematically may be of several types. In whole-language programs, the emphasis is upon meaning-based reading and writing activities. Phonics instruction is integrated into these activities but taught incidentally as teachers decide it is needed. Basal programs consist of a teacher's manual and a complete set of books and materials that guide the teaching of beginning reading. Some basal programs focus on whole-word or meaning-based activities with limited attention to letter-sound constituents of words and little or no instruction in how to blend letters to pronounce words. In sight word programs, children begin by building a reading vocabulary of 50 to 100 words, and then later they learn about the alphabetic system. These types of non-phonics programs were among those taught to children in the control groups of experiments examined by the NRP. Distinctions among the various types of non-phonics programs are not absolute. However, their defining characteristic is that they do not provide explicit, systematic phonics instruction.

Phonics programs have been used to teach young children to read as they progress through the primary grades and to remediate the reading difficulties of poor readers. The Panel analyzed studies that examined the effectiveness of phonics programs with three types of problem readers: children in kindergarten or 1st grade who were at risk for developing reading problems; older children of average or better intelligence who were not making normal progress in reading, referred to as disabled readers; older children who were progressing poorly in reading and who varied in intelligence with at least some of them achieving poorly in other academic areas, referred to as low-achieving readers.

For children to learn to read, several capabilities must be developed. The focus of systematic phonics instruction is on helping children acquire knowledge of the alphabetic system and its use to decode new words, and to recognize familiar words accurately and automatically. Knowing how letters correspond to phonemes and larger subunits of words is essential for enabling beginning readers to sound out word segments and blend these parts to form recognizable words. Alphabetic knowledge is needed to figure out new words by analogy and to help beginners remember words they have read before. Knowing letter-sound relations also helps children to be more accurate in predicting words from context. In short, knowledge of the alphabetic system contributes greatly to children's ability to read words in isolation or connected text.

To study whether systematic phonics instruction improves children's ability to read words in various ways, different measures have been used. Decoding was tested by having children read regularly spelled words. To test whether children could read novel words, pseudowords (e.g., gan, bloff, trusk) were used. Sight vocabulary was examined through sets of leveled, miscellaneous words, not all of which were spelled regularly. In addition to word-reading, children's performance on measures of oral reading, text comprehension, and spelling was measured.

To provide solid evidence, experiments to test the contribution of systematic phonics instruction to reading acquisition must be well designed. Random assignment of students to treatment and control groups is a procedure that controls for other factors and allows researchers to conclude that the treatment itself was the cause of any growth in reading. However, sometimes the realities of schools and teachers make it impossible to randomly assign students, so researchers have to use quasi-experimental designs, assigning treatment and control conditions to already existing groups. Although researchers should administer pretests to determine whether the treatment and control groups differed prior to treatment and then remove any differences statistically when outcomes are analyzed, this is not always done. Also, larger sample sizes provide more reliable findings, but access to many students is not always possible. In evaluating the evidence, the Panel attempted to rule out weak designs as the explanation for any positive effects that were produced by systematic phonics instruction.

Methodology

To evaluate the evidence, the NRP conducted a meta-analysis. The literature was searched electronically to locate potential studies. To qualify for the analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group.
2. Studies had to appear in a refereed journal after 1970.
3. Studies had to provide data testing the hypothesis that systematic phonics instruction improves reading performance more than instruction providing unsystematic phonics or no phonics instruction. To be considered an instance of phonics instruction, the treatment had to teach children to identify or use symbol-sound correspondences systematically.
4. Studies had to measure reading as an outcome.
5. Studies had to report statistics permitting the calculation or estimation of effect sizes.
6. Studies were not those already included in the NRP's meta-analysis of phonemic awareness training studies.

From the potentially relevant list of references, 75 studies that appeared to meet the criteria were identified and located. These were carefully reviewed to determine their suitability for the meta-analysis. Studies of instructional interventions that might be found in schools were sought. Short-term laboratory studies and studies that taught only a limited set of processes were eliminated. Also eliminated were studies that simply compared different forms of phonics instruction but did not include a control group receiving reduced phonics or no phonics. Of the 75 studies screened, 38 were retained and 37 were eliminated from the final set used to calculate effect sizes.

The primary statistic used in the analysis of performance on outcome measures was effect size, indicating whether and by how much performance of the treatment group exceeded performance of the control group, with the difference expressed in standard deviation units. From the 38 studies entered into the database, 66 treatment-control group comparisons were derived.

Studies were coded for several characteristics that were included as moderators in the meta-analysis:

- Type of phonics program (synthetic programs emphasizing instruction in the sounding out and blending of words vs. programs teaching students to decode using larger subunits of words such as phonograms, as well as letters and sounds vs. miscellaneous programs),
- Specific phonics programs that were evaluated in at least three different studies (Direct Instruction; Lippincott; Orton Gillingham; Sing Spell Read and Write; Benchmark Word ID; New Primary Grades Reading System)
- Type of program taught to the control group (basal program, regular curriculum, whole language approach, whole word program, miscellaneous programs)
- Group assignment procedure (random assignment or nonequivalent groups)
- Number of participants (blocked into quartiles)
- Grade level (kindergarten, 1st grade, 2nd through 6th grades)
- Reading ability (normally developing, at risk, low achiever, reading disabled)
- Socioeconomic status (low, middle, varied, not given)
- Instructional delivery unit (class, small groups, 1:1 tutoring).

Children identified as being low achieving or at risk for reading failure were those tested and shown to have poor letter knowledge, poor phonemic awareness, or poor reading skills, or those in schools with low achievement, or those identified by teachers as needing special help in reading, or those who qualified for remedial programs in schools but the criteria for selection were not specified. Children classified as reading disabled were those identified according to IQ-reading discrepancy criteria in standard use by researchers or those given tests to determine that the disability was reading-specific. In some cases, exclusionary criteria were applied as well (e.g., no neurological, behavioral, or emotional disorders).

Across the studies, the effects of phonics instruction on reading were most commonly assessed at the end of training. For programs lasting longer than one year, outcomes were measured at the end of each year in most cases. The primary outcome used in the meta-analysis was that assessed at the end of training or at the end of one year, whichever came first. Effect sizes were calculated on six types of outcome measures:

- Decoding regularly spelled real words
- Reading novel words in the form of pseudowords
- Reading miscellaneous words some of which were irregularly spelled
- Spelling words
- Comprehending text read silently or orally
- Reading text accurately aloud.

The mean effect size across these measures was calculated to yield a general literacy measure for each comparison. A statistical program was employed to calculate effect sizes and to test the influence of moderator variables on effect sizes. An effect size of $\underline{d} = 0.20$ is considered small; a moderate effect size is $\underline{d} = 0.50$; an effect size of $\underline{d} = 0.80$ or above is large.

Results and Conclusions

There were 38 studies from which 66 treatment-control group comparisons were derived. Although each comparison could contribute up to six effect sizes, one per outcome measure, few studies did. The majority (76%) of the effect sizes involved reading or spelling single words while 24% involved text reading. The imbalance favoring single words is not surprising given that the focus of phonics instruction is on improving children's ability to read and spell words. Moroever, many of the studies were conducted with beginning readers whose reading development at the time of the study was too limited to assess textual reading. Studies limiting instructional attention to children with reading problems accounted for 65% of the comparisons, 38% involving poor readers considered at risk or low achieving, and 27% diagnosed as reading disabled (RD). Studies involving first graders were overrepresented in the database, accounting for 38% of the comparisons. Fewer kindergartners (12%) and children in 2nd through 6th grades (23%) were represented. Children in the RD group spanned several ages and grades, ranging from ages 6 to 13 and grades 2 through 6. Most of the studies (72%) were recent, conducted in the last 10 years.

Systematic phonics instruction typically involves explicitly teaching students a prespecified set of letter-sound relations and having students read text that provides practice using these relations to decode words. Instruction lacking an emphasis on phonics instruction does not teach letter-sound relations systematically and selects text for children according to other principles. The latter form of instruction includes whole word programs, whole language programs, and some basal reader programs.

The meta-analyses were conducted to answer several questions about the impact of systematic phonics instruction on growth in reading when compared to instruction that does not emphasize phonics. Findings provided strong evidence substantiating the impact of systematic phonics instruction on learning to read.

1. Does systematic phonics instruction help children learn to read more effectively than nonsystematic phonics instruction or instruction teaching no phonics?

Children's reading was measured at the end of training if it lasted less than a year or at the end of the first school year of instruction. The mean overall effect size produced by phonics instruction was moderate in size and statistically greater than zero, $\underline{d} = 0.44$. Findings provided solid support for the conclusion that systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs providing unsystematic or no phonics instruction.

2. Are some types of phonics instruction more effective than others? Are some specific phonics programs more effective than others?

Three types of phonics programs were compared in the analysis: (1) synthetic phonics programs which emphasized teaching students to convert letters (graphemes) into sounds (phonemes) and then to blend the sounds to form recognizable words; (2) larger-unit phonics programs which emphasized the analysis and blending of larger subparts of words (i.e., onsets, rimes, phonograms, spelling patterns) as well as phonemes; (3) miscellaneous phonics programs that taught phonics systematically but did this in other ways not covered by the synthetic or larger-unit categories or were unclear about the nature of the approach. The analysis showed that effect sizes for the three categories of programs were all significantly greater than zero and did not differ statistically from each other. The effect size for synthetic programs was d = 0.45, for larger-unit programs, d = 0.34, and for miscellaneous programs, d = 0.27. The conclusion supported by these findings is that various types of systematic phonics approaches are significantly more effective than non-phonics approaches in promoting substantial growth in reading.

There were seven programs that were examined in three or more treatment-control group comparisons in the database. Analysis of the effect sizes produced by these programs revealed that all were statistically greater than zero and none differed statistically from the others in magnitude. Effect sizes ranged from d = 0.23 to 0.68. In most cases there were only three or four comparisons contributing effect sizes, so results may be unreliable. The conclusion drawn is that specific systematic phonics programs are all significantly more effective than non-phonics programs; however, they do not appear to differ significantly from each other in their effectiveness although more evidence is needed to verify the reliability of effect sizes for each program.

3. Is phonics taught more effectively when students are tutored individually or when they are taught in small groups or when they are taught as classes?

All three delivery systems proved to be effective ways of teaching phonics, with effect sizes of d = 0.57 (tutoring), d = 0.43 (small group), and d = 0.39 (whole class). All effect sizes were statistically greater than zero, and no one differed significantly from the others. This supports the conclusion that systematic phonics instruction is effective when delivered through tutoring, through small groups, and through teaching classes of students.

4. Is phonics instruction more effective when it is introduced to students not yet reading, in kindergarten or 1st grade, than when it is introduced in grades above 1st after students have already begun to read?

Phonics instruction taught early proved much more effective than phonics instruction introduced after first grade. Mean effect sizes were kindergarten d = 0.56; first grade d = 0.54; 2nd through 6th grades d = 0.27. The conclusion drawn is that phonics instruction produces the biggest impact on growth in reading when it begins in kindergarten or 1st grade before children have learned to read independently. These results indicate clearly that systematic phonics instruction in kindergarten and 1st grade is highly beneficial and that children at these developmental levels are quite capable of learning phonemic and phonics concepts. To be effective, systematic phonics instruction introduced in kindergarten must be appropriately designed for learners and must begin with foundational knowledge involving letters and phonemic awareness.

5. Is phonics instruction beneficial for children who are having difficulty learning to read? Is it effective in preventing reading failure among children who are at risk for developing reading problems in the future? Is it effective in remediating reading difficulties in children who have been diagnosed as reading disabled and children who are low-achieving readers?

Phonics instruction produced substantial reading growth among younger children at risk of developing future reading problems. Effect sizes were d = 0.58 for kindergartners at risk and d = 0.74 for 1st graders at risk. Phonics instruction also significantly improved the reading performance of disabled readers (i.e., children with average IQs but poor reading) for whom the effect size was d = 0.32. These effect sizes were all statistically greater than zero. However, phonics instruction failed to exert a significant impact on the reading performance of low-achieving readers in 2nd through 6th grades (i.e., children with reading difficulties and possibly other cognitive difficulties explaining their low achievement). The effect size was d = 0.15, which was not statistically greater than chance. Possible reasons might be that the phonics instruction provided to low-achieving readers was not sufficiently intense, or that their reading difficulties arose from sources not treated by phonics instruction such as poor comprehension, or there were too few cases (i.e., only eight treatment-control comparisons pulled from three studies) to yield reliable findings.

The conclusion drawn from these findings is that systematic phonics instruction is significantly more effective than non-phonics instruction in helping to prevent reading difficulties among at risk students and in helping to remediate reading difficulties in disabled readers. No conclusion is drawn in the case of low-achieving readers because it is unclear why systematic phonics instruction produced little growth in their reading and whether the finding is even reliable. Further research is needed to determine what constitutes adequate remedial instruction for low-achieving readers.

6. Does phonics instruction improve children's reading comprehension ability as well as their decoding and word-reading skills?

Systematic phonics instruction was most effective in improving children's ability to decode regularly spelled words (d = 0.67) and pseudowords (d = 0.60). This was expected because the central focus of systematic phonics programs is upon teaching children to apply the alphabetic system to read novel words. Systematic phonics programs also produced growth in the ability to read irregularly spelled words although the effect size was significantly lower, d = 0.40. This is not surprising because a decoding strategy is less helpful for reading these words. However, alphabetic knowledge is useful for establishing connections in memory that help children read irregular words they have read before. This may explain the contribution of phonics.

Systematic phonics instruction produced significantly greater growth than non-phonics instruction in younger children's reading comprehension ability (d = 0.51). However, the effects of systematic phonics instruction on text comprehension in readers above 1st grade were mixed. Although gains were significant for the subgroup of disabled readers (d = 0.32), they were not significant for the older group in general (d = 0.12).

The conclusion drawn is that growth in word-reading skills is strongly enhanced by systematic phonics instruction when compared to non-phonics instruction for kindergartners and 1st graders as well as for older struggling readers. Growth in reading comprehension is also boosted by systematic phonics instruction for younger students and reading disabled students. These findings should dispel the any belief that teaching phonics systematically to young children interferes with their ability to read and comprehend text. Quite the opposite is the case. Whether growth in reading comprehension is produced generally in students above 1st grade is less clear.

7. Does phonics instruction have an impact on children's growth in spelling?

Systematic phonics instruction produced much growth in spelling among the younger students, that is, kindergartners and 1st graders, d = 0.67, but not among the older students (above 1st grade), whose effect size of d = 0.09 did not differ significantly from zero. One factor contributing to the difference is that younger children were given credit for using phonics-based knowledge to produce letter-sound spellings of words as well correct spellings whereas older children were not. Another factor may be that as children move up in the grades, remembering how to spell words requires knowledge of higher level regularities not covered in phonics programs. A third reason for the poor showing among older students may be that the majority were poor readers, known to have difficulty learning to spell.

The conclusion drawn is that systematic phonics instruction contributed more than non-phonics instruction in helping kindergartners and 1st graders apply their knowledge of the alphabetic system to spell words. However, it did not improve spelling in students above 1st grade.

8. Is phonics instruction effective with children at different SES levels?

Systematic phonics instruction helped children at all SES levels make significantly greater gains in reading than did non-phonics instruction. The effect size for low SES students was d = 0.66 and for middle-class students was d = 0.44. Both were statistically greater than zero and did not differ from each other. The conclusion drawn is that systematic phonics instruction is beneficial to students regardless of their SES.

9. Does the type of control group used to evaluate the effectiveness of phonics instruction make a difference?

The type of nonsystematic or non-phonics instruction given to control groups to evaluate the effectiveness of systematic phonics instruction varied across studies and included the following types: basal programs, regular curriculum, whole language approaches, whole word programs, and miscellaneous programs. The question of whether systematic phonics instruction produced better reading growth than each type of control group was answered affirmatively in each case. The effect sizes were all positive favoring systematic phonics, were all statistically greater than zero, and ranged from d = 0.31 to 0.51. No single effect size differed from any of the others.

The conclusion supported by these findings is that the effectiveness of systematic phonics instruction found in the present meta-analysis did not depend on the type of instruction that students in the control groups received. Students taught phonics systematically outperformed students who were taught a variety of nonsystematic or non-phonics programs, including basal programs, whole language approaches, and whole-word programs.

10. Were studies reporting the largest effects of phonics instruction well designed or poorly designed experiments? That is, was random assignment used? Were the sample sizes sufficiently large? Might results be explained by differences between treatment and control groups that existed prior to the experiment rather than by differences produced by the experimental intervention?

The effects of systematic phonics instruction were not diminished when only the best designed experiments were singled out. The mean effect size for studies using random assignment to place students in treatment and control groups, d = 0.45, was essentially the same as that for studies employing quasi-experimental designs, d = 0.43, which used existing groups to compare phonics instruction and non-phonics instruction. The mean effect size for studies administering systematic phonics and non-phonics instruction to large samples of students did not differ from studies using the fewest students. For studies using between 80 and 320 students, d = 0.49; for studies using between 20 and 31students, d = 0.48. There were some studies that did not use random assignment and either failed to address the issue of pre-existing differences between treatment and control groups or mentioned that a difference existed but did not adjust for differences in their analysis of results. The effect sizes changed very little when these comparisons were removed from the database, from d = 0.44 to d = 0.46.

The conclusion drawn is that the significant effects produced by systematic phonics instruction on children's growth in reading were evident in the most rigorously designed experiments. Significant effects did not arise primarily from the weakest studies.

11. Is enough known about systematic phonics instruction to make recommendations for classroom implementation? If so, what cautions should be kept in mind by teachers implementing phonics instruction?

Findings of the Panel regarding the effectiveness of systematic phonics instruction were derived from studies conducted in many classrooms with typical classroom teachers and typical American or English-speaking students from a variety of backgrounds and SES levels. Thus, the results of the analysis are indicative of what can be accomplished when systematic phonics programs are implemented in today's classrooms. Systematic phonics instruction has been used widely over a long period of time with positive results. A variety of phonics programs have proven effective with children of different ages, abilities, and socioeconomic backgrounds. These facts should persuade educators and the public that systematic phonics instruction is a valuable part of a successful classroom reading program. The Panel's findings summarized above serve to illuminate the conditions that make phonics instruction especially effective. However, caution is needed in giving a blanket endorsement to all kinds of phonics instruction.

It is important to recognize that the goals of phonics instruction are to provide children with some key knowledge and skills and to insure that they know how to apply this knowledge in their reading and writing. Phonics teaching is a means to an end. To be able to make use of letter-sound information, children need phonemic awareness. That is, they need to be able to blend sounds together to decode words, and they need to break spoken words into their constituent sounds to write words. Programs that focus too much on the teaching of letter-sounds relations and not enough on putting them to use are unlikely to be very effective. In implementing systematic phonics instruction, educators must keep the *end* in mind and insure that children understand the purpose of learning letter-sounds and are able to apply their skills in their daily reading and writing activities.

In addition to this general caution, several particular concerns should be taken into consideration to avoid misapplication of the findings. One concern relates to the commonly heard call for "intensive, systematic" phonics instruction. Usually the term "intensive" is not defined, so it is not clear how much teaching is required to be considered "intensive." Questions needing further answers are: How many months or years should a phonics program continue? If phonics has been taught systematically in kindergarten and 1st grade, should it continue to be emphasized in 2nd grade and beyond? How long should single instructional sessions last? How much ground should be covered in a program? That is, how many letter-sound relations should be taught and how many different ways of using these relations to read and write words should be practiced for the benefits of phonics to be maximum? These are among the many questions that remain for future research.

Secondly, the role of the teacher needs to be better understood. Some of the phonics programs showing large effect sizes are scripted in such a way that teacher judgment is largely eliminated. Although scripts may standardize instruction, they may reduce teachers' interest in the teaching process or their motivation to teach phonics. Thus, one concern is how to maintain consistency of instruction and at the same time encourage unique contributions from teachers. Another concern involves what teachers need to know. Some phonics programs require a sophisticated understanding of spelling, structural linguistics, and word etymology. Teachers who are handed the programs but are not provided with sufficient inservice training to use these programs effectively may become frustrated. In view of the evidence showing the effectiveness of systematic phonics instruction, it is important to ensure that the issue of how best to prepare teachers to carry out this teaching effectively and creatively is given high priority. Knowing that all phonics programs are not the same brings with it the implication that teachers must themselves be educated about how to evaluate different programs, to determine which are based on strong evidence and how they can most effectively use these programs in their own classrooms.

As with any instructional program, there is always the question: "Does one size fit all?" Teachers may be expected to use a particular phonics program with their class, yet it quickly becomes apparent that the program suits some students better than others. In the early grades, children are known to vary greatly in the skills they bring to school. There will be some children who already know most letter-sound correspondences, some children who can even decode words, and others who have little or no letter knowledge. Should teachers proceed through the program and ignore these students? Or should they assess their students' needs and select the types and amounts of phonics suited to those needs? Although the latter is clearly preferable, this requires phonics programs that provide guidance in how to place students into flexible instructional groups and how to pace instruction. However, it is common for many phonics programs to present a fixed sequence of lessons scheduled from the beginning to the end of the school year.

Finally, it is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program. Phonics instruction is never a total reading program. In 1st grade, teachers can provide controlled vocabulary texts that allow students to practice decoding, and they can also read quality literature to students to build a sense of story and to develop vocabulary and comprehension. Phonics should not become the dominant component in a reading program, neither in the amount of time devoted to it nor in the significance attached. It is important to evaluate children's reading competence in many ways, not only by their phonics skills but also by their interest in books and their ability to understand information that is read to them. By emphasizing all of the processes that contribute to growth in reading, teachers will have the best chance of making every child a reader.

Directions for Further Research

Although phonics instruction has been the subject of a great deal of study, there are important topics that have received little or no research attention, and there are other topics that, although previously studied, require further research to refine our understanding.

Three important but neglected questions are prime candidates for research: What are the "active ingredients" in effective systematic phonics programs? Is phonics instruction improved when motivational factors are taken into account — not only learners' but also teachers' motivation to teach? How does the use of decodable text as early reading material contribute to the effectiveness of phonics programs?

1. Active Ingredients

Systematic phonics programs vary in many respects. It is important to determine whether some properties are essential and others are not. Because instructional time during the school day is limited, teachers and publishers of beginning reading programs need to know which ingredients of phonics programs yield the most benefit.

2. Motivation

Phonics instruction has often been portrayed as involving "dull drill" and "meaningless worksheets." Few if any studies have investigated the contribution of motivation to the effectiveness of phonics programs, not only the learner's motivation to learn but also the teacher's motivation to teach. The lack of attention to motivational factors by researchers in the design of phonics programs is potentially very serious because debates about reading instruction often boil down to concerns about the "relevance" and "interest value" of *how* something is being taught, rather than the specific content of *what* is being taught. Future research on phonics instruction should investigate how best to motivate children in classrooms to learn the letter-sound associations and to apply that knowledge to reading and writing. It should also be designed to determine which approaches teachers prefer to use and are most likely to use effectively in their classroom instruction.

3. Decodable Text

Some systematic phonics programs are designed so that children are taught letter-sound correspondences and then provided with little books written carefully to contain the letter-sound relations that were taught. Some programs begin with a very limited set and expand these gradually. The intent of providing books that match children's letter-sound knowledge is to enable them to experience success in decoding words that follow the patterns they know. The stories in such books often involve pigs doing jigs and cats in hats. Systematic phonics programs vary in the percentage of decodable words in 1st-grade stories and in the percentage of sight words introduced holistically to make a good story. Surprisingly, very little research has attempted to determine the contribution of decodable books to the effectiveness of phonics programs.

There are other important topics to be addressed in future research as well. These include the following:

- Should systematic phonics instruction continue beyond 2nd grade? If so, what are the goals of more advanced forms of phonics instruction and does this instruction contribute to growth in reading?
- Are there ways to improve the effectiveness of systematic phonics instruction for poor readers above 1st grade? Does this instruction need to take account of any maladaptive reading habits the students have acquired or any sources impeding the incorporation of alphabetic knowledge and decoding strategies into their reading? Does this instruction need to take account of the type of reading instruction they experienced in earlier years? Does decoding instruction need to be combined with comprehension instruction?

PART II: PHONICS INSTRUCTION

Report

Introduction

Learning to read is a complex task for beginners. They must coordinate many cognitive processes to read accurately and fluently. Readers must be able to apply their alphabetic knowledge to decode unfamiliar words and to remember how to read words they have read before. When reading connected text, they must construct sentence meanings and retain them in memory as they move on to new sentences. At the same time, they must monitor their word recognition to make sure that the words activated in their minds fit with the meaning of the context. In addition, they must link new information to what they have already read, as well as to their background knowledge, and use this to anticipate forthcoming information. When one stops to take stock of all the processes that readers perform when they read and comprehend text, one is reminded how amazing the act of reading is and how much there is for beginners to learn.

In teaching phonics explicitly and systematically, several different instructional approaches have been used. These include synthetic phonics, analytic phonics, embedded phonics, analogy phonics, onset-rime phonics, and phonics through spelling. Although these explicit and systematic phonics approaches all use a planned, sequential introduction of a set of phonic elements with teaching and practice of those elements, they differ across a number of other features. For example, the content covered ranges from a limited to an elaborate set of letter-sound correspondences and phonic generalizations. The application procedures taught to children vary. Synthetic phonics programs teach children to convert letters into sounds or phonemes and then blend the sounds to form recognizable words. Analytic phonics avoids having children pronounce sounds in isolation to figure out words. Rather, children are taught to analyze letter-sound relations once the word is identified. Phonics-through-spelling programs teach children to transform sounds into letters to write words. Phonics in context approaches teach children to use sound-letter correspondences along with context cues to identify unfamiliar words they encounter in text. Analogy phonics programs teach children to use parts of written words they already know to identify new words. The distinctions between systematic phonics approaches are not absolute, however, and some phonics programs combine two or more of these types of instruction. In addition, these approaches differ with respect to the extent that controlled vocabulary (decodable text) is used for practicing reading connected text. Although these differences exist, the hallmark of systematic phonics programs is that they delineate a planned, sequential set of phonic elements, and they teach these elements, explicitly and systematically. The goal is to enable learners to acquire sufficient knowledge and use of the alphabetic code so that they can make normal progress in learning to read and comprehend written language.

A key feature that distinguishes systematic phonics instruction from nonsystematic phonics is in the identification of a full array of letter-sound correspondences to be taught. The array includes not only the major correspondences between consonant letters and sounds but also short and long vowel letters and sounds, and vowel and consonant digraphs (e.g., oi, ea, ou, sh, ch, th). Also, it may include blends of letter-sounds that recur as subunits in many words, such as initial blends (e.g., st, sm, bl, pr), and final stems (e.g., -ack, -end, -ill, -op). Learning vowel and digraph spelling patterns is harder for children: therefore, special attention is devoted to learning these relations. It is not sufficient just to teach the alphabetic system. Children need practice in applying this knowledge in reading and writing activities. Programs provide practice in various ways. Phonics programs may teach children decoding strategies that involve sounding out and blending individual letters and digraphs, or pronouncing and blending larger subunits such as initial blends and final stems of words. Programs may provide children with text whose words can be decoded using the letter-sound relations already taught. Programs may have children write their own text using the letter-sounds taught and then have children read their own and others' stories.

The purpose of literacy instruction in schools is to help children master the many challenges of written language. While teachers use a variety of activities to accomplish this purpose, one central approach is to teach the alphabetic code that represents oral language in writing. Children need to understand how letters, called graphemes, stand for the smallest sounds, called phonemes, in spoken words. Systematic phonics instruction teaches beginning readers the alphabetic code consisting of a large set of correspondences between graphemes and phonemes and perhaps larger sub-units of words and how to use this knowledge to read words. In some phonics programs, beginners are taught a routine for transforming spellings into blends of phonemes that are recognized as words. Learning about letter-sound associations helps beginners break the code in learning to read. However, the English writing system has other higher level, word-based regularities as well, so, although phonics instruction contributes, it is not the complete solution to word identification that it is in other written languages that are more fully phonemic (e.g., Spanish).

Over the years educators have disagreed about how beginning reading should be taught. Some have advocated starting with a systematic phonics approach while others have argued for a whole word approach or a whole language approach. Disagreement has centered on whether teaching should begin with systematic explicit instruction in symbol-sound correspondences, whether it should begin with whole words, or whether initial instruction should be meaning-centered with correspondences taught incidentally in context as needed. Most recently the pendulum has swung toward providing children with more explicit phonics instruction. Educators advocating this shift have claimed that there is substantial research showing that approaches with an emphasis on phonics instruction are more effective than approaches that do not emphasize the teaching of phonics.

The purpose of this report was to examine the research evidence concerning phonics instruction. The Panel sought answers to the following questions:

- Does systematic phonics instruction help children learn to read more effectively than unsystematic phonics instruction or instruction teaching no phonics?
- Are some types of phonics instruction more effective than others? Are some specific phonics programs more effective than others?
- Is phonics instruction more effective when it is introduced to students not yet reading, in kindergarten or 1st grade, than when it is introduced in grades above 1st after students have already begun to read?
- Is phonics instruction beneficial for children who are having difficulty learning to read? Is it effective in preventing reading failure among children who are at risk for developing reading problems in the future? Is it effective in remediating reading difficulties among children who have not made normal progress in learning to read?
- Is phonics taught more effectively when students are tutored individually, or when they are taught in small groups, or when they are taught as classes?
- Does phonics instruction improve children's ability to read connected text as well as their decoding and word reading skills?
- Does phonics instruction have an impact on children's growth in spelling?
- Is phonics instruction effective with children at different socioeconomic levels?
- Does the type of instruction given to control groups and used to evaluate the effectiveness of phonics instruction make a difference? That is, is systematic phonics more effective than forms of instruction that do not emphasize phonics, such as the whole word approach or meaning-centered approaches?
- If phonics instruction is found to be more effective than less-phonics or no-phonics instruction, were the experiments showing these effects well designed or poorly designed?

To evaluate the evidence, a meta-analysis was conducted. The Panel searched the literature to locate experimental studies published after 1970 that administered systematic phonics instruction to one group of children and administered another type of instruction that involved unsystematic phonics or no phonics to a control group. Also the studies had to examine phonics programs of the sort used in schools rather than single-process-focused laboratory procedures. The studies had to measure reading as an outcome of instruction. In addition, studies were excluded if they were in the Panel's other database used to conduct a meta-analysis examining effects of phonemic awareness instruction on reading. A total of 38 studies meeting the NRP research criteria was found. The studies were coded for various characteristics of students, instruction, and experimental design. A meta-analysis was conducted to examine the size of effects that resulted when the performance of students receiving systematic phonics instruction was compared to that of students receiving another form of instruction that did not focus on phonics. The outcomes measured following instruction included children's ability to read words and pseudowords, to read and comprehend text, and also to spell words.

Background and Rationale for the Meta-Analysis

Historical Overview

The question of whether instruction that includes an initial emphasis on systematic phonics is more effective than other forms of instruction in teaching children to read has been addressed many times in the literature. The particular issues underlying interest in this question have shifted over the years, but the topic has remained controversial, and this has spawned a number of reviews of research.

In the 1960s, the Office of Education funded the Cooperative Research Program in First Grade Reading (Bond & Dykstra, 1967, 1998) and Project Literacy (Levin & Williams, 1970). The First Grade studies involved a wide-ranging research project, consisting of 29 separate studies in different sites, all aimed at determining the "best" approach to teaching beginning reading. In contrast, Project Literacy attempted to identify the basic psychological and linguistic processes involved in learning to read and did not focus directly on the pedagogy of reading. At the same time, the Carnegie Foundation funded Jeanne Chall's (1967) comprehensive review of beginning reading instruction, Learning to Read: The Great Debate. That review, like the present report, was intended to analyze the results of previous research.

Concern about beginning reading instruction was not confined just to the educational community but was very much in public discourse. Flesch (1955) had authored a best selling book Why Johnny Can't Read in which he argued that children were being abused by the then-current whole word methodology. Flesh asserted that if children were taught only the 44 letter-sound correspondences, they would be able to read any word they encountered, and there would be no reading problems. Spurred on partially by Flesch and partially by advances in linguistics, new phonics programs were developed and began achieving wider usage in reading instruction (Aukerman, 1981; Popp, 1975).

Chall's (1967) review examined both the underlying theory and the classroom realities of these new phonics programs. But the core of her study was a comprehensive analysis of the research up to the mid-1960s, including the then-unpublished First Grade Studies. Chall's basic conclusion continues to be cited to this day, her finding that early and systematic instruction in phonics seems to lead to better achievement in reading than later and less systematic phonics instruction.

It is important to note that Chall, in the 1967 edition of her review, did not recommend any particular type of phonics instruction. Common forms of phonics instruction in the 1960s included synthetic instruction, analytic instruction, and linguistic readers (Aukerman, 1981). All of these challenged the sight word approach of the day. However, in the 1983 edition of her review, Chall did suggest that synthetic phonics instruction held a slight edge over analytic phonics instruction. Even in this, her recommendation was temperate.

Chall's (1967) basic finding has been reaffirmed in nearly every research review conducted since then (e.g., Adams, 1990; Anderson et al., 1985; Balmuth, 1982). Also, one of the coordinators of the First Grade Studies (Dykstra, 1968) published an analysis in which he concluded that the results of that project supported Chall's basic finding (Adams, 1990). Nevertheless, the controversy has persisted over this issue (Grundin, 1994; Taylor, 1998; Weaver, 1998). Part of the reason that the debate has continued is that phonics instruction has become entangled with politics and ideology (Goodman, 1993; McKenna, Stahl, & Reinking, 1994; Stahl, 1999). Another reason has been philosophical disagreements about how children learn to read and confusions about the implications of these varied points of view.

Phonics and No-Phonics Instruction

At the time of Chall's (1967) original review, the contrast between phonics and the alternative "look-say" methods was considerable. In the look-say approach, children were taught to read words as wholes much like Chinese logographs, and they practiced reading words until they had acquired perhaps 50 to 100 words in their sight vocabularies. Only after this accomplishment, which occurred toward the end of 1st grade, did phonics instruction begin. This was truly non-phonics instruction because discussion of letter-sound relations was delayed for a considerable length of time. The look-say approach contrasted with a variety of phonics programs. These included synthetic phonics programs which taught children to sound out and blend words, linguistic programs which taught decoding through patterned words and phonetically controlled texts, and analytic phonics programs which taught children to analyze letter-sound relations in previously learned words so as to avoid pronouncing sounds in isolation (Aukerman, 1971, 1984).

In the present day, whole language approaches have replaced the whole word method as the alternative to systematic phonics programs. The shift has involved a change from very little letter-sound instruction in 1st grade to a modicum of letter-sounds taught unsystematically. In contrast to the whole word method, whole language teachers are not told to wait until a certain point before teaching children about letter-sound relationships. Whereas in the 1960s, it would have been easy to find a 1st grade reading program without any phonics instruction, in the 1980s and 1990s this would be rare. Baumann, Hoffman, Moon, and Duffy-Hester (1998), in a national survey of 1,207 elementary school teachers, found that 63% believed that phonics should be taught directly and that 89% believed that skills instruction should be combined with literature and language-rich activities. Fisher, Lapp, and Flood (1999), in a survey of 118 California teachers, found that 64% of the K through 2 teachers integrated phonics instruction into their lessons (with some extra isolated phonics), and the remainder taught phonics as a separate part of word study.

Whole language teachers typically provide some instruction in phonics, usually as part of invented spelling activities or through the use of graphophonemic prompts during reading (Routman, 1996). However, their approach is to teach it unsystematically and incidentally in context as the need arises. The whole language approach regards letter-sound correspondences, referred to as graphophonemics, as just one of three cueing systems (the others being semantic/meaning cues and syntactic/language cues) that are used to read and write text. Whole language teachers believe that phonics instruction should be integrated into meaningful reading, writing, listening, and speaking activities and taught incidentally when they perceive it is needed. As children attempt to use written language for communication, they will discover naturally that they need to know about letter-sound relationships and how letters function in reading and writing. When this need becomes evident, teachers are expected to respond by providing the instruction.

Although some phonics is included in whole language instruction, important differences have been observed distinguishing this approach from systematic phonics approaches. In several vignettes portraying phonics instruction in whole language contexts (Dahl, Sharer, Lawson, & Grogran, 1999; Freppon & Dahl, 1991; Freppon & Headings, 1996; Mills, O'Keefe, & Stephens, 1992), few if any instances of vowel instruction were found (Stahl, Duffy-Hester, & Stahl, 1998). This contrasts with systematic phonics programs where the teaching of vowels is central and is considered essential for enabling children to decode (Shankweiler & Liberman, 1972).

Another practice that is found in some systematic phonics programs but is not found in whole language programs is that of teaching children to say the sounds of letters and blend them to decode unfamiliar words. Programs that teach this procedure are referred to as synthetic phonics programs. Systematic phonics programs also commonly teach children an extensive, pre-specified set of letter-sound correspondences or phonograms while whole language programs teach a more limited set, in context, as needed. Systematic phonics programs teach phonics explicitly by delineating a planned, sequential set of phonic elements and teaching these elements explicitly and systematically; some systematic phonics programs also use controlled vocabulary (decodable text) to provide practice with these elements. Whole language programs do not prespecify the relations to be taught. It is presumed that exposing children to letter-sound relations as they read text will foster incidental learning of the relations they need to develop as readers.

The meta-analysis was conducted to compare the effectiveness of systematic phonics instruction to other forms of instruction lacking an emphasis on phonics. Included in the database were several studies that provided whole language instruction to control groups and studies teaching whole word programs to control groups. In fact, two studies in the database were conducted for the purpose of evaluating the effects of whole language programs, not phonics programs. In these studies, phonics was the form of instruction given to control groups (Klesius et al., 1991; Freppon, 1991).

Not only whole language and whole word instruction but also other forms of control-group instruction were present in the database. Several control groups received some type of basal instruction, usually a program prescribed by the school or district. Basal programs consist of a whole package of books and supplementary materials that are used to teach reading. Teachers work from a thick manual that details daily lesson plans based on a scope and sequence of the reading skills to be taught. Students are given workbooks to practice on skills. Tests are used to place students in the proper levels of the program and to assess mastery of skills (Aukerman, 1981). Basal reading programs do vary, but one can assume that basal readers of the same era are roughly similar in their characteristics. The basal programs given to control groups provided only limited or no systematic phonics instruction.

A few studies utilized as their baseline control the performance of comparable classes of students enrolled in the same schools the year prior to the treatment (Snider, 1990; Vickery et al., 1987). In one case, a basal program was used. In the other case, the type of program was not specified. Campbell and Stanley (1966) suggest that this design contains certain threats to external validity, especially the differential history of the two groups.

Some studies in the database included more than one control group. The Panel selected for the meta-analysis the group receiving the least phonics instruction.

The issue of the control group is crucial. A meta-analysis compares a treatment to what is supposedly a constant. However, in reality, the size of the effect is a result of what goes on in both the treatment and the control groups. A treatment can be very effective but yield only a small effect size if instruction in the control group is also effective. On the other hand, if the control group's instruction is particularly ineffective, by design or by accident, then the effect size is inflated. One must consider the nature of the control group in order to interpret an effect size. The question addressed in the meta-analysis was whether phonics instruction produced greater growth in reading than each of the various types of instruction given to control groups.

Types of Phonics Instruction

The hallmarks of systematic phonics programs are that children receive explicit, systematic instruction in a set of prespecified associations between letters and sounds, and they are taught how to use them to read, typically in texts containing controlled vocabulary. However, phonics programs vary considerably in exactly what children are taught and how they are taught (Adams, 1990; Aukerman, 1981). Approaches to phonics instruction may differ in several important ways including the following:

1. How many letter-sound relations are taught, how they are sequenced, whether phonics generalizations are taught as well (e.g., "When there are two vowels side by side, the long sound of the first one is heard and the second is usually silent."), whether special marks are added to letters to indicate their sounds, for example, curved or straight lines above vowels to mark them as short or long
2. The size of the unit taught (i.e., graphemes and phonemes, or larger word segments called phonograms, for example, -ing, or -ack which represent the rimes in many single-syllable words)
3. Whether the sounds associated with letters are pronounced in isolation (synthetic phonics) or only in the context of words (analytic phonics)
4. The amount and type of phonemic awareness that is taught, for example, blending or segmenting sounds orally in words

5. Whether instruction is sequenced according to a hierarchical view of learning with the steps regarded as a series of prerequisites (i.e., letters, then letter-sound relations, then words, then sentences) or whether multiple skills are learned together
6. The pace of instruction
7. The word reading operations that children are taught, for example, sounding out and blending letters, or using larger letter subunits to read words by analogy to known words
8. The involvement of spelling instruction
9. Whether learning activities include extensive oral drill-and-practice, reciting phonics rules, or filling out worksheets
10. The type of vocabulary control provided in text (e.g., is the vocabulary limited mainly to words containing familiar letter-sound associations or are sight words introduced to help create a meaningful story?)
11. Whether phonics instruction is embedded in or segregated from the literacy curriculum
12. The teaching approach, whether it involves direct instruction in which the teacher takes an active role and students passively respond, or whether a "constructivist" approach is used in which the children learn how the letter-sound system works through problemsolving
13. How interesting and motivating the instructional activities are for teachers and for students.

Systematic phonics programs included in the Panel's database varied in many of these ways; so, it should not be assumed that the programs taught phonics uniformly. One purpose of the meta-analysis was to examine whether different properties of phonics programs influenced how effective they were in teaching children to read. However, this purpose was thwarted by the fact that most studies did not describe the phonics instruction in sufficient detail to permit coding the properties listed above. As a result, the Panel selected only one property for coding: whether programs emphasized a synthetic approach in teaching children to read words or whether the emphasis was on larger subunits of words.

A majority of the programs in the database used a synthetic approach to teach phonics. This instruction typically begins by teaching children relations between individual letters and pairs of letters called digraphs (e.g., TH, AI, CH, OI) and all 44 sounds or phonemes of the language. These correspondences are introduced systematically and sequentially. Children are taught to decode unfamiliar words by sounding out the letters and blending them to pronounce a recognizable word.

However, the synthetic strategy presents two difficulties for children. One is that blending words containing stop consonants requires deleting "extra" (schwa vowel) sounds produced when letters are pronounced separately, for example, blending "tuh-a-puh" requires deleting the "uh" sounds to produce the blend "tap." The second problem is that when the sounds to be blended exceed two or three, it becomes harder to remember and manage the ordering of all those sounds, for example, blending "s-tuh-r-ea-m" to say "stream."

Phonics programs have been developed to address these difficulties. One approach used has been to teach students to read larger subunits of words as well as phonemes. For example, children learn to recognize ST, AP, EAM, as blends so that there are not so many separate parts of words to sound out and remember in blending them. The larger units taught might include onsets (i.e., the consonants that precede the vowel such as "st" in stop) and rimes (i.e., the vowel and following consonants such as "op" in stop), also called phonograms, and spelling patterns characterizing the common parts of word families (e.g., -ack as in pack and stack, -oat as in goat and float). Teaching children to analyze and pronounce parts of words provides the basis for teaching them the strategy of reading new words by analogy to known words (e.g., reading stump by analogy to jump). In the database, these studies are distinguished and classified as teaching children to analyze and blend words by using larger phonological units.

The database included 43 treatment-control comparisons that taught synthetic phonics to the treatment groups, 11 studies that used phonics treatments emphasizing larger subunits for blending words, two comparisons that combined both types of programs, and ten comparisons that fit neither category, referred to as miscellaneous. In the meta-analysis, effect sizes of the three larger sets of phonics types were compared.

In the database were seven phonics programs whose effectiveness was assessed in at least three different treatment-control group comparisons. All but one of the programs, Lovett's analogy program, taught synthetic phonics. These programs together with the dates of publication are listed below:

- Direct Instruction, also referred to as DISTAR and Reading Mastery (1969, 1978, 1979, 1980, 1987, 1988)
- Lovett's adaptation of Direct Instruction (1994)
- Lovett's adaptation of the Benchmark Word Identification program (1994)
- The Lippincott Basic Reading program (1963, 1981)
- Beck and Mitroff's New Primary Grades Reading System (1972)
- Orton Gillingham programs (1940, 1956, 1969, 1979, 1984)
- Sing, Spell, Read, and Write (1972).

For each program, there were at least three treatment-control group comparisons testing effects of that form of phonics instruction; so, effect sizes were examined separately in a meta-analysis. Most of these programs were developed over 20 years ago, providing researchers with more time to study them than recently developed programs. The question addressed in the meta-analysis was whether these programs were effective in promoting growth in reading and whether they differed in effectiveness. There was no apriori reason to expect any differences. Likewise there was no reason to expect these programs to be more effective than programs not in the set being compared.

Grade and Reading Ability

A question of particular interest to the Panel was when should phonics instruction begin. Should it be introduced in kindergarten when children may know very little about letters, phonemic awareness, or should it be started in 1st grade after children have received prereading or emergent reading experiences in kindergarten? According to Chall (1996a, b), beginners need to develop foundational knowledge such as concepts about print, phonological awareness, and letter names prior to formal reading instruction. Studies indicate that knowing letters and having phonemic awareness are essential for learning to use the alphabetic system to read and spell words (see the NRP review of phonemic awareness instruction). Thus, formal, systematic phonics instruction that expects students to learn to decode words in kindergarten may be too much.

On the other hand, in countries such as New Zealand and the United Kingdom, the practice of introducing children to reading and writing at the age of 5 in full-day programs has existed for many years. The Reading Recovery® program (Clay, 1993) is designed to pick up the stragglers having difficulty at the age of 6, when North American children are typically just beginning reading instruction. Thus, the notion that kindergartners are not ready for formal reading instruction at age 5 is questionable.

In some studies in the database, a middle road was taken. Children were introduced to simplified reading and spelling activities using a basic set of letters and sounds that they were taught. Instruction began by providing a foundation for students and then building on this to ease students into reading when they became ready for it. (See Blachman et al., 1999; Vandervelden & Siegel, 1997). In the meta-analysis, the contribution of phonics instruction at the kindergarten level was examined across studies that varied in how much phonics material was covered.

The most important grade for teaching phonics is thought to be 1st grade when formal instruction in reading typically begins in the United States. Children have foundational knowledge and are ready to put it to use in learning to read and write. In contrast, introducing phonics instruction in grades above 1st means that children who were taught to read in some other way may be required to switch gears in order to incorporate phonics procedures into their reading and writing. The database included studies that introduced phonics to students at various grade levels. The question addressed in the meta-analysis was whether the grade level in which phonics instruction was introduced made any difference in the outcomes observed. Another related question is whether phonics instruction that was started in kindergarten is more effective than phonics instruction begun in 1st or 2nd grade. Data were probed for an answer to this question as well.

Phonics instruction has also been widely regarded as particularly beneficial to children with reading problems (e.g., Foorman et al., 1998). Many studies have shown that reading disabled children have exceptional difficulty decoding words (Rack, Snowling, & Olson, 1992). In fact, their level of performance falls below that of younger non-disabled readers who read at the same grade-equivalent level, indicating a serious deficit in decoding skill. Phonics instruction that teaches disabled readers to decode words should remediate this deficit and should enable these students to make better progress in learning to read. The meta-analysis evaluated the contribution made by phonics instruction to growth in reading among children having difficulty learning to read.

Two types of children with reading problems have been distinguished by researchers, children who are *unexpectedly* poor readers because their intelligence (an index of learning aptitude for some academic skills) is *higher* than their reading ability, and children whose below-average reading is not surprising given that their intelligence is also below average. Various labels such as dyslexic or learning disabled or reading disabled have been applied to children whose higher IQs are discrepant with their poor reading skill. Children whose lower reading scores are consistent with their lower IQs have been referred to as low achievers or garden variety poor readers (Stanovich, 1986). The question of interest was whether phonics instruction helps to remediate reading difficulties for both types of poor readers. Studies in the database were brought to bear on this question.

Delivery Systems for Teaching Phonics

There are various delivery systems that might be used to teach phonics. Tutoring one-on-one is regarded as the ideal form of instruction for students who are having difficulties because it allows teachers to tailor lessons to address individual students' needs. One of the best known tutoring programs is Reading Recovery® (Clay, 1993). The database included three studies that modified Reading Recovery® lessons to include systematic phonics instruction (Greaney et al., 1997; Santa & Hoien, 1999; Tunmer & Hoover, 1993). A total of eight studies taught phonics through tutoring. The remainder of the studies utilized small groups or whole classes to deliver instruction. Of interest was whether one type of delivery system produced greater gains in reading than the other types. In the Panel's analysis of phonemic awareness training effects, comparison of instructional units revealed that small groups produced superior learning. However, it was expected that tutoring would be the most effective way to teach phonics.

Word Reading Processes: Assessing Growth

It is important to distinguish between the *methods* of teaching reading and the *processes* that learners acquire as they receive instruction and learn to read. Sometimes the two may be confused. For example, the term "sight word" has a "methods" meaning and a "process" meaning. As a method, sight words are the high-frequency, irregularly spelled words students are taught to read as unanalyzed wholes, often on flash cards, for example, said, once, their, come. In contrast, the "process meaning" of sight words refers to words that are stored in readers' heads and that enable them to read those words immediately upon seeing them. Not just high-frequency words but all words that readers practice reading become retained as sight words in memory.

Methods of teaching reading are aimed at helping learners acquire the processes they need to develop skill as readers. In considering how phonics instruction promotes growth in reading, it is important to describe the reading processes that learners are expected to acquire.

Learning to read can be analyzed as involving two basic processes (Gough & Tunmer, 1986; Hoover & Gough, 1990). One process involves learning to convert the letters into recognizable words. The other involves comprehending the meaning of the print. When children attain reading skill, they learn to perform both of these processes so that their attention and thought are focused on the meaning of the text while word reading processes operate unobtrusively and out of awareness for the most part. Children acquire comprehension skill in the course of learning to speak. Comprehension processes that children use to understand spoken language are thought to be the same ones that they use to read and understand text. In contrast, children do not acquire word reading skill in the course of learning to speak. This achievement requires special experiences and instruction.

Many mental processes are active when readers read and understand text. Readers draw on their knowledge of language to create sentences out of word sequences. They access their background knowledge to construct meaning from the text. They retain this information in memory and update it as they interpret more text. Readers monitor their comprehension to verify that the information makes sense.

A central part of text processing involves reading the words. Four different ways can be distinguished (Ehri, 1991, 1994):

1. Decoding: Readers convert letters into sounds and blend them to form recognizable words; the letters might be individual letters, or digraphs such as TH, SH, OI, or phonograms such as ER, IGH, OW, or spellings of common rimes such as -AP, -OT, -ICK. Ability to convert letter subunits into sounds comes from readers' knowledge of the alphabetic system.
2. Sight: Readers retrieve words they have already learned to read from memory.
3. Analogy: Readers access in memory words they have already learned and use parts of the spellings to read new words having the same spellings (e.g., using -ottle in bottle to read throttle).
4. Prediction: Readers use context cues, their linguistic and background knowledge, and memory for the text to anticipate or guess the identities of unknown words.

Text reading is easiest when readers have learned to read most of the words in the text automatically by sight because little attention or effort is required to process the words. When written words are unfamiliar, readers may decode them or read them by analogy or predict the words, but these steps take added time and shift attention at least momentarily from the meaning of text to figuring out the words.

Readers need to learn how to read words in the various ways to develop reading skill. The primary way to build a sight vocabulary is to apply decoding or analogizing strategies to read unfamiliar words. These ways of reading words help the words to become familiar. Processing letter-sound relations in the words through decoding or analogizing creates alphabetic connections that establish the words in memory as sight words (Ehri, 1992; Share, 1995).

Systematic phonics instruction is thought to contribute to the process of learning to read words in these various ways by teaching readers use of the alphabetic system. Alphabetic knowledge is needed to decode words, to retain sight words in memory, and to call on sight word memory to read words by analogy. In addition, the process of predicting words from context benefits from alphabetic knowledge. Word prediction is made more accurate when readers can combine context cues with letter-sound cues in guessing unfamiliar words in text (Tunmer & Chapman, 1998).

One purpose of the meta-analysis was to examine whether phonics instruction improves readers' ability to decode words and to read words by sight. To study the impact of phonics instruction on the various ways to read words, different measures have been used. The ability to decode words is tested by giving children regularly spelled words to read. The ability to decode novel words never read before is tested by having children read pseudowords. Children's sight vocabulary is examined by giving them miscellaneous words including irregularly spelled words that are ordered by grade level from preprimer to the highest grades.

Methodology

Database

An electronic search was conducted in two databases, ERIC and PsycINFO. Three sets of terms were used in the search. These terms were derived by the Panel on the basis of analyses of various reference guides including the Literacy Dictionary (Harris & Hodges, 1995), the Handbook of Research on Teaching the English Language Arts (Flood, Jensen, Lapp, & Squire, 1991), the Encyclopedia of English Studies and the Language Arts (Purves, 1994), and the Handbook of Reading Research (Barr, Kamil, Mosenthal, & Pearson, 1991; Pearson, Barr, Kamil, & Mosenthal, 1984).

- **Set 1:** Alphabetic code, analogy approach, code emphasis, compare-contrast, decodable text, decoding, phonemic decoding, phonetic decoding, phonological decoding, direct code, direct instruction, Reading Mastery, explicit instruction, explicit phonological processes, grapheme-phoneme correspondences, graphophonic, Initial Teaching Alphabet, letter training, letter-sound correspondences, linguistic method, McCracken, Orton-Gillingham, phoneme analysis, phoneme blending, phoneme-grapheme correspondences, phonics, Alphabetic phonics, analytic phonics, embedded phonics, structured phonics, synthetic phonics, systematic phonics, phonological processing, Recipe for Reading, recoding, phonological recoding, Slingerland approach, Spaulding approach, word study, word sort, words by analogy. These were combined using "or" statements, meaning that all articles indexed by any of these terms would be located.

- **Set 2:** Beginning reading, beginning reading instruction, instruction, intervention, learning to decode, reading improvement, reading instruction, remedial training, remedial reading, remediation, teaching, training, disabled readers, dyslexia, reading difficulties, reading disability, reading failure, reading problems. These were combined in the search using "or" statements.
- **Set 3:** Miscues, oral reading, reading ability, reading achievement, reading acquisition, reading aloud, reading comprehension, reading development, reading processes, reading skills, silent reading, story reading, word attack, word identification, word recognition, word reading, nonword reading. These, too, were combined with "or" statements.

The three sets of terms were used to locate potentially relevant studies in the two databases. Articles selected were those that included at least one term from each set. Because the term spelling had not been included in Set 1, the search was run a second time with spelling crossed with Set 2 and Set 3 terms. The first search uncovered 391 articles in PsycINFO and 520 articles in ERIC. The second search uncovered 252 articles in PsycINFO and 210 articles in ERIC. Abstracts were printed and screened.

To qualify for the analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group.
2. Studies had to appear in a refereed journal after 1970.
3. Studies had to provide data testing the hypothesis that systematic phonics instruction improves reading performance more than instruction providing unsystematic phonics or no phonics instruction. To be considered an instance of phonics instruction, the treatment had to teach children to identify or use symbol-sound correspondences systematically.
4. Studies had to measure reading as an outcome.
5. Studies had to report statistics permitting the calculation or estimation of effect sizes.
6. Studies were not those already included in the National Reading Panel's meta-analysis of phonemic awareness training studies.

From the various lists of references, 75 studies that appeared to meet the criteria were identified and located. The goal was to analyze studies that resembled each other so that the corpus would be more homogeneous. Studies of instructional interventions that might be found in schools were sought. Short-term laboratory studies and studies that provided instruction on only a limited set of processes were eliminated. Also eliminated were studies that simply compared different forms of phonics instruction but did not include a control group receiving reduced phonics or no phonics. Of the 75 studies screened, 38 were retained and 37 were eliminated from the final set used to calculate effect sizes. The reasons for eliminating studies and the numbers of studies eliminated are listed in Table 1 on the next page.

Some minor deviations from the above procedures occurred. More recent studies that would not yet have appeared in electronic searches were obtained from current issues of journals and preprints of in press papers sent to members of the Panel. Also, Blachman et al. (1999) conducted a 3-year longitudinal study to evaluate the effects of phonemic awareness and phonics instruction on children as they progressed from kindergarten through 2nd grade. Results of the first year were published as a separate study and included in the Panel's phonemic awareness meta-analysis. Results of the more extensive 3-year study were included in the phonics instruction database. This was the only study analyzed in both reports.

Table 1

Reasons for Excluding Studies From the Database

BASIS FOR REJECTION	NUMBER
Control group missing or inadequate:	5 studies
Short-term, focused too limited, or laboratory study:	14 studies
Inadequate statistics:	8 studies
Inadequate outcome measures:	3 studies
Not a study of phonics instruction:	2 studies
Duplicate data reported in another publication already considered:	5 studies
Total:	37 studies

The primary statistic used in the analysis of performance on outcome measures was effect size, indicating whether and by how much performance of the treatment group exceeded performance of the control group, with the difference expressed in standard deviation units. The formula used to calculate raw effect sizes for each treatment-control comparison consisted of the mean of the treatment group minus the mean of the control group divided by a pooled standard deviation.

From the 38 studies entered into the database, 66 treatment-control group comparisons were derived. There were six cases in which the same control group was compared to two different phonics treatment groups. There was one study in which the same control group was compared to four different treatments (Lovett et al., in press). Each comparison was treated as a separate case with separate effect sizes in the database.

Studies were coded for several characteristics that were included as moderators in the meta-analysis:

- Type of phonics program (synthetic vs. larger subunits vs. a combination of synthetic and larger subunits vs. miscellaneous)
- Specific phonics program if replicated in at least three comparisons
- Type of control group (basal, regular instruction, whole language, whole word, miscellaneous)
- Group assignment procedure (random assignment or nonequivalent groups)
- Number of participants (blocked into quartiles)
- Grade level or age
- Reading ability (normally developing, at risk/low achiever, reading disabled)
- Socioeconomic status (low, middle, varied, not given)
- Instructional delivery unit (class, small groups, 1:1 tutoring).

The studies, their properties, and effect sizes are listed in Appendix G.

Although the length of treatment was coded, it was not used as a moderator variable. Many of the studies were vague about the amount of time devoted to phonics instruction; so, it was not possible to calculate precise amounts of time spent, particularly in classroom studies which provided instruction regularly throughout the school year. Also, treatment length was confounded with other variables considered to be more important, such as whether students were tutored or taught in classes, whether students were poor or normally developing readers, whether students were beginners or older readers when they began instruction.

Some studies in the database selected normally developing readers to include in their experiments whereas other studies singled out poor readers. These students were grouped into four types of readers for analysis:

1. Normally developing readers: this category included studies in which poor readers were excluded and studies where no attempt was made to distinguish children by reading ability.
2. Disabled readers: this category included children who were identified as reading disabled according to IQ-reading discrepancy criteria in standard use by researchers, or were given tests to determine that the disability was reading-specific; in some cases, exclusionary criteria were applied as well (e.g., no neurological, behavioral, economic, or emotional disorders); most of these children were above 1st grade.
3. Children at risk for developing reading difficulties in the future (kindergartners and 1st graders).
4. Children who were below average in their reading referred to as low achievers (children above 1st grade).

The latter two groups included children who exhibited poor letter knowledge, poor phonemic awareness, or poor reading skills, or those in schools with low achievement, or those identified by teachers as needing special help in reading, or those who qualified for remedial programs in schools but the criteria for selection were not specified. The at-risk label was applied to children in kindergarten and 1st grade because they were still at a beginning level in their learning. Children labeled low achievers in reading were those in 2nd grade and above whose identity as poor readers was considered to be better established. Both groups included children who also had lower than average IQs qualifying them as garden variety poor readers with generally low academic achievement, but the groups were not limited to children with low IQs because researchers either did not measure IQ or did not use it to limit the readers selected for study.

Six types of outcomes assessing growth in reading or spelling were distinguished:

- Decoding of real words chosen to contain regular spelling-to-sound relationships
- Reading nonsense words or pseudowords chosen to represent regular spelling-to-sound relationships.
- Word identification (in some cases, words were chosen to represent irregular spelling-to-sound relationships)
- Spelling, assessed using either developmental stages for younger children (Bear et al., 2000) or number of words correct
- Comprehension of material read silently or orally
- Oral reading of connected text (accuracy).

Measures reported in studies were classified into these types, and effect sizes were computed for each type of outcome. Some studies included several measures of an outcome type and reported means on each measure. In these cases, effect sizes were calculated on each measure and then averaged. This step insured that no single treatment-control comparison contributed more than one effect size to any single outcome category. Some studies included tests to assess whether students were able to read or spell words that were taught directly during phonics instruction. These results were not included as outcomes in the database.

For each comparison, the mean effect size was calculated across whichever of the six measures had been assessed in that study. This yielded an overall outcome measure for each comparison. When studies reported performance on a general reading test but no more specific tests, the overall effect size was based on the general measure. Outcomes that did not fit into the above categories were not entered into the database.

Performance of students was measured at various points before, during, and after instruction. Entered into the database were outcomes of posttests measured at three points in time: at the end of training, at the end of the first school year if the program was taught for more than one year, after a delay following training to assess long-term effects. The type of posttest most commonly given was that occurring at the end of the program or at the end of the school year when the program continued; so, this was the outcome used in most of the analyses of moderator variables.

In the categorization of outcome measures, no distinction was drawn between standardized and experimenter-devised tests. Comprehension measures tended to be standardized. Oral reading measures tended to be informal reading inventories that were neither standardized nor developed specifically for the study. Word lists were both standardized and experimenter-devised. Standardized tests of word reading most commonly came from the Woodcock Johnson Achievement series, the Woodcock Reading Mastery Test, and the Wide Range Achievement (WRAT) test. In general, standardized measures tend to produce smaller effect sizes than experimenter-devised measures. This was observed in the NRP's analysis of effects of phonemic awareness instruction on measures of word reading and spelling. One reason is that standardized tests are designed to assess reading across a wide range of ability levels and hence are less sensitive to differences at any one level in the range. Thus, aggregating the two types of tests would be expected to underestimate effect sizes slightly.

The information and statistics required to generate and analyze effect sizes were entered into a separate database using Microsoft Excel and SPSS. The data entered included identification of the study, codes for the information listed above, means and standard deviations of treatment and control groups on outcome measures, pooled standard deviations, raw effect sizes (g) and effect sizes weighted for the size of the sample (d). When means and standard deviations were not available in the article, DSTAT was used to estimate effect sizes based on t or F values. When pretest differences between treatment and control groups were reported, effect sizes were calculated to eliminate these differences as far as possible.

The DSTAT statistical package (Johnson, 1989) was employed to calculate effect sizes and to test the influence of moderator variables on effect sizes. Each moderator variable had at least two levels. Tests were conducted to determine whether the mean weighted effect size (d) at each level was significantly greater than zero at $p < 0.05$, whether the individual effect sizes at each level were homogeneous ($p < 0.05$), and whether effect sizes differed significantly at different levels of the moderator variables ($p < 0.05$).

Consistency With the Methodology of the National Reading Panel

The methodology approved by the National Reading Panel was adopted. The search was conducted in accordance with most of the prescribed procedures. Studies that were not published in peer-reviewed journals were excluded. All of the studies in the data base utilized experimental or quasi-experimental designs. (Studies using a multiple baseline design were not included.) The studies were coded for most of the specified categories plus some additional categories of interest for this particular analysis. Properties left uncoded were those where information was rarely provided. More properties were coded than were considered in the analysis. One reason for not analyzing effects of moderator (coded) variables on outcomes was that there were insufficient numbers of comparisons to provide a valid analysis of these effects.

The Panel determined that a meaningful meta-analysis could be conducted on the data. The means and standard deviations that were used to calculate effect sizes were verified by checking all of them at least twice. Intercoder reliability was conducted on the variables used in the meta-analysis and exceeded the prescribed level of 90%. Disagreements were resolved by discussion and consensus.

Results

Characteristics of Studies in the Data Set

There were 38 studies from which 66 treatment-control group comparisons were derived. Each comparison could contribute a maximum of six effect sizes, one per outcome measure. However, few studies included measures of all the outcomes. The most commonly assessed outcome (i.e., at the end of training or at the end of one year, whichever came first) was word identification consisting of 59 effect sizes. The least common outcome was oral reading with 16 effect sizes. The other outcomes ranged from 30 to 40 effect sizes. Whereas 76% of the effect sizes involved reading or spelling single words, only 24% involved text reading. Although there is a marked imbalance favoring single words, this is not surprising given that phonics instruction is aimed primarily at improving children's ability to read and spell words.

Many of the studies limited instructional attention to children with reading problems. These studies accounted for 65% of the comparisons, with 38% involving poor readers considered "at risk" or low achieving, and 27% involving children diagnosed as reading disabled (RD). Studies involving 1st graders were overrepresented in the database compared to other grades and accounted for 38% of the comparisons. Fewer studies involved kindergartners and children in 2nd through 6th grades, with these groups contributing 12% and 23% of the comparisons, respectively. Children in the RD group spanned several ages and grades, ranging from ages 6 to 13 and grades 2 to 6. Several properties of the studies in our database were examined. Of interest was whether the studies were older or more recent. A tally revealed the following distribution:

1970 to 1979: 1 study

1980 to 1989: 9 studies

1990 to 2000: 28 studies

Thus, the majority of the studies were conducted over the last 10 years. Most (66%) were carried out in the United States, but 24% were done in Canada, and the remainder in the United Kingdom, Australia, and New Zealand. Thus, the evidence came from a variety of locales. Other properties of comparisons in the database are listed in Table 2 in Appendix D.

Effects of Phonics Instruction on Outcome Measures

The statistic used to assess the effectiveness of phonics instruction on children's growth in reading was effect size which measures how much the mean of the phonics group exceeded the mean of the control group in standard deviation units. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, suggesting a strong effect of training. An effect size of 0 indicates that treatment and control group means were identical, suggesting that training had no effect. To judge the strength of an effect size, values suggested by Cohen (1988) are commonly used. An effect size of 0.20 is considered small; a moderate effect size is 0.50; an effect size of 0.80 or above is large.

An overall effect size was calculated for each of the 66 treatment-control group comparisons. This was the average of the six specific outcome effect sizes (i.e., decoding, word reading, comprehension, etc.) or the effect size from a general reading measure if no specific outcomes were measured. In the analyses, this overall effect size is interpreted as assessing the impact of phonics instruction on growth in reading. Although one of the six was a spelling measure, spelling effect sizes contributed only 16% of the effect sizes that were averaged and reading measures contributed the rest (84%). Mean effect sizes obtained on various outcomes associated with levels of the moderator variables are reported in Table 3 (Appendix E). Effect sizes were tested statistically to determine whether each was significantly greater than zero, indicating that superior performance of phonics-trained groups over control groups was not a result of chance at $\underline{p} < 0.05$.

Inspection across the effect sizes listed in Table 3 reveals that the vast majority were significantly greater than zero (those marked with an asterisk). This indicates that systematic phonics instruction was effective across a variety of conditions and characteristics. The overall mean effect size of phonics instruction on reading was $\underline{d} = 0.41$ when effects of programs were tested at their conclusion. A few programs lasted longer than 1 school year. To obtain another index of effects, outcomes measured either at the end of the program or the end of the first school year, whichever came first, were calculated. Results revealed an effect size of $\underline{d} = 0.44$. These findings indicate that the effect produced by phonics instruction on reading was moderate in size. Unless otherwise stated, the test point used to assess effects of moderator variables in the meta-analyses was that occurring at the end of training or at the end of the first school year, whichever came first.

Phonics instruction in most of the studies lasted 1 school year or less. However, there were four treatment-control comparisons in which longer training was provided. In these studies, children at risk for reading problems began phonics instruction in kindergarten or 1st grade and continued for 2 or 3 years. Outcomes were measured at the end of each school year (Blachman et al., 1999; Brown & Felton, 1990; Torgesen et al., 1999). Characteristics and results of the four comparisons drawn from these studies are presented in Table 4. Mean effect sizes across the four comparisons were sizeable and their strength was maintained across the grades: kindergarten d = 0.46; 1st grade d = 0.54; 2nd grade d = 0.43. This indicates the value of starting phonics early and continuing to teach it for 2 to 3 years. (See results below for additional evidence regarding the value of teaching phonics early.) In the Blachman et al. (1999) study, instruction was not given to all 2nd graders but only to those who had not attained the goals of the program after 2 years of instruction. These findings point to the importance of programs providing tests for teachers to use to determine which children need additional systematic phonics instruction and which have mastered the processes taught.

A few studies examined effects of phonics instruction several months after the treatment had ended. The specific comparisons together with their properties are listed in Table 4 (Appendix E). Followup tests were administered from 4 months to 1 year after training. As shown in Table 3, the effect size remained significantly greater than zero, indicating that the impact of phonics instruction lasted well beyond the end of training although its size was somewhat diminished (from d = 0.51 to d = 0.27).

The aim of phonics instruction is to help children acquire knowledge and use of the alphabetic system to read and spell words. Phonics was expected to exert its greatest impact on the ability to decode regularly spelled words and nonwords. Phonics instruction was also expected to exert a large effect when spelling was measured using a developmental spelling scale, which gives credit for letter-sound spellings as well as correct spellings (e.g., Bear et al., 2000; Blachman et al., 1999). These capabilities all benefit directly from alphabetic knowledge. Phonics instruction was expected to exert a significant but smaller impact on the ability to read miscellaneous words that included irregularly spelled words. Although alphabetic knowledge is not helpful for decoding irregularly spelled words, it does help children remember how to read these words (Ehri, 1998). Phonics instruction was expected to impact text reading processes. The effect was expected to be significant but smaller because its influence is indirect.

From Table 3 (Appendix E), it is apparent that effect sizes for all six types of measures were statistically greater than zero, indicating that phonics instruction significantly improved performance on all of the outcome measures examined, not only word reading and spelling but also text processing. Inspection of the size of the effects provided support for the various hypotheses. The strongest effects occurred on measures of decoding regularly spelled words (d = 0.67) and pseudowords (d = 0.60). These effects were statistically larger than effects observed on the other measures which did not differ from each other. This indicates that phonics instruction was especially effective in teaching children to decode novel words, one of the main goals of phonics.

Effect sizes on comprehension measures (d = 0.27) and oral reading measures (d = 0.25) were statistically greater than zero, indicating that phonics instruction significantly improved children's text processing skills as well as their word reading skills. The fact that effects of phonics instruction on reading comprehension were positive serves to dispel any belief that teaching phonics to children interferes with their ability to read and comprehend text. Quite the opposite is the case.

Several reasons explain why effects were somewhat smaller on text processing measures than on word reading measures. The tests of comprehension were predominantly standardized tests which are less sensitive when the range of performance is limited. The target of phonics instruction is teaching children how to read words. Although word recognition skill influences how well children can read and comprehend text, there are other processes that are important as well. Moreover, readers can still get meaning from text even when they cannot read some of the words.

Analysis of Moderator Variables

Studies in the database varied in several respects that were coded and analyzed as moderator variables. Of interest was whether these moderator variables enhanced or limited the effectiveness of systematic phonics instruction on growth in reading. It is important to recognize the limitations of this type of analysis and the tentative nature of any conclusions that are drawn. Findings involving the impact of moderator variables on effect sizes cannot support strong claims about moderators being the cause of the difference. Moderator findings are no more than correlational. The biggest source of uncertainty is whether there is a hidden variable that is confounded with the moderator and is the true cause of the difference.

Characteristics of Students

The students who received phonics instruction across the studies varied in two important ways that were expected to make a difference on the effect sizes produced by phonics instruction: their age or grade in school, and their reading ability. Kindergartners, particularly those at risk, know little about letters and sounds. Typically they are nonreaders. For them, phonics instruction begins by teaching letter shapes, letter sounds, phonemic awareness, and how to apply these in simplified reading and writing tasks. Later in kindergarten or at the beginning of 1st grade, formal reading instruction begins with much ground to cover. Children typically start as emergent readers and by the end of 1st grade are able to read text independently. In systematic phonics programs, extensive instruction is provided to develop children's knowledge of the alphabetic system and how to use this knowledge to read words in and out of text. The greatest impact of phonics instruction is expected to occur in helping 1st graders get off the ground in learning to read.

Designers of phonics programs to teach beginning reading expect children to start receiving instruction in their programs when the children are in kindergarten or 1st grade before they have acquired any reading skill. Programs are designed so that children usually continue receiving instruction at least through 2nd grade. What happens when these programs are taught to children above 1st grade who have already acquired some reading skill with some other program is less clear. Are the older children given 1st grade catch-up instruction? Do the phonics strategies that they are taught compete or conflict with the reading skills and strategies that they have already acquired? If so, what is done about this instructionally? There are many uncertainties surrounding the introduction of phonics instruction to children in the upper grades who have already moved into reading.

The database that the Panel analyzed included several studies with older children beyond 1st grade. Many of these studies involved disabled readers or low achieving readers who received remedial instruction designed to address the problems of poor readers. However, there were also a few studies in which phonics instruction was provided to normally developing readers who had already received instruction in other unspecified programs in the earlier grades. It is important to recognize that the question addressed in the meta-analysis of these studies was whether introducing phonics instruction presumably as a new program for these older children was effective in promoting their growth in reading.

Younger vs. Older Children

To analyze the impact of age and grade combined, two groups of children were distinguished: the younger children in kindergarten and 1st grade; and the older students in 2nd through 6th grades. The latter group included the mixed age/grade comparisons involving reading disabled (RD) children and low achieving readers. The outcome variable was the effect sizes on the immediate posttest given either at the end of training or at the end of the first year of the program, whichever came first.

From Table 3 (Appendix E), it is apparent that systematic phonics instruction produced a significant impact on children's growth as readers in both groups, as indicated by effect sizes statistically greater than zero. However, phonics instruction made a larger contribution to younger children's growth as readers ($d = 0.55$) than to older children's growth ($d = 0.27$). The difference in effect sizes favoring younger children was statistically significant.

The pool of effect sizes among the younger students was not homogeneous; so, effects were examined separately for kindergartners and 1st graders. From Table 2, it is evident that effect sizes were very similar, $d = 0.56$ for kindergartners and $d = 0.54$ for 1st graders. This shows that a moderate and significant effect size typified children in both grades. According to Chall (1992), phonics instruction should exert its greatest impact in the early grades. These findings show that effects were equally strong in both kindergarten and 1st grade, indicating that "early" includes both of these grades. There were many more studies of the impact of phonics in 1st grade than in kindergarten, so the 1st grade findings are more reliable than the kindergarten findings.

Whereas the database on phonics instruction included only seven comparisons involving kindergartners, the National Reading Panel's database of phonemic awareness training studies included 40 kindergarten comparisons that measured reading as an outcome. In the PA analysis, effects were moderate in size and statistically significant. The effect size in the PA analysis (d = 0.48) was close to the effect size produced by phonics instruction (d = 0.58). Combined, these findings clearly support the importance of teaching phonemic awareness and grade-appropriate phonics in kindergarten. Indeed, some of the phonemic awareness training studies that taught children to analyze phonemes using letters would have qualified as phonics studies. If these PA studies had not been excluded from the phonics database, there would have been more kindergarten comparisons.

The above findings suggest that when phonics instruction is introduced and taught in kindergarten or 1st grade to readers who have little reading ability, it produces a larger effect than when phonics is introduced in grades above 1st grade with readers who have already acquired some reading skills. However, before concluding that phonics is truly less effective with older children, it is important to consider several mitigating factors. The majority of the comparisons in the older group, 78%, involved either low achieving or disabled readers. Remediating their reading problems may be especially difficult. In addition, there were only seven comparisons involving older, normally developing readers, and four of these came from one study using the Orton-Gillingham method, a program developed for disabled readers, not for non-disabled upper elementary level readers. Perhaps other types of phonics programs designed expressly to improve reading in older non-disabled children might prove more effective. This question awaits more research.

The set of effect sizes for the older students proved to be homogeneous, indicating that chance, rather than other moderator variables, explains the variation in effect sizes. The two types of poor readers, low achievers and RDs, contributed the majority of the effect sizes to this pool. These findings indicate that low achieving readers and disabled readers do not differ in their response to phonics instruction.

Specific Outcomes in Younger Readers

Because the younger and older children differed in their response to phonics instruction, the question of whether phonics instruction impacted children's ability to decode and spell words and to read text was answered separately for the two groups. Results in Table 3 (Appendix E) show that, among kindergartners and 1st graders, phonics instruction produced significant growth on all six outcome measures whose effect sizes were statistically greater than zero. Because a central goal in phonics programs is to teach students to decode novel words, one would expect the strongest effects to be evident in decoding tasks. This is what was found. The largest effect size was produced on the measure of decoding regularly spelled words (d = 0.98). Moderately large effects were also produced on measures of decoding pseudowords (d = 0.67) and spelling words (d = 0.67). The effect size was somewhat reduced on the word identification outcome (d = 0.45). This is not surprising since tests of word identification often included irregularly spelled words not amenable to decoding.

Phonics instruction with its emphasis on teaching letter-sound relations would be expected to improve beginning readers' ability to spell words by writing the sounds they hear. Studies with younger children commonly employed developmental spelling scoring systems that gave credit for phonetically plausible spellings, for example, spelling feet as FET or car as KR (Tangel & Blachman, 1995; Morris & Perney, 1984). This may explain the sizeable effect observed on the spelling outcome (d = 0.67).

Among beginning readers, phonics instruction exerted a significant impact on reading comprehension. The effect size, based on ten 1st grade and one kindergarten comparisons, was moderate (d = 0.51). However, the effect size on another measure of text reading, oral reading, was smaller but also significantly greater than zero (d = 0.23 based on two kindergarten and four 1st grade comparisons). Why phonics skills facilitated reading comprehension more than oral reading is not clear. It may have to do with the nature of the tests. Standardized comprehension tests at this level generally use extremely short (usually one sentence) "passages." On these short passages, the effects of decoding should be strong. Some tests, such as the Gates-MacGinitie, favor phonetically regular words in these passages. Oral reading measures, on the other hand, use longer passages, sometimes containing pictures which would enhance the utility of context.

One would expect effect sizes on text reading and word reading to be similar because 1st graders' ability to read and understand text is heavily influenced by their ability to read the words in the text, perhaps somewhat more so than in later grades. This is supported by Juel (1994) who found a very high correlation between word recognition and reading comprehension in 1st grade (r = 0.87) and found that the correlation was somewhat lower in 2nd grade (r = 0.73).

In sum, these findings show that systematic phonics instruction helped beginning readers acquire and use the alphabetic system to read and spell words in and out of text. Children who were taught phonics systematically benefited significantly more than beginners who did not receive phonics instruction in their ability to decode regularly spelled words and nonwords, in their ability to remember how to read irregularly spelled words, and in their ability to invent phonetically plausible spellings of words. In addition, phonics instruction contributed substantially to students' growth in reading comprehension and somewhat less to their oral text reading skill.

Specific Outcomes in Older Readers

Students above the 1st grade were introduced to phonics instruction in their classes or in pull-out programs for periods lasting up to a school year. These students included children who were low achieving readers as well as children diagnosed as reading disabled. Effects of phonics instruction on six outcome measures were compared. Results in Table 3 (Appendix E) show that substantial growth occurred in learning to decode regularly spelled words (d = 0.49) and pseudowords (d = 0.52), with effect sizes statistically greater than zero in the moderate range. This shows that phonics programs were significantly more effective than control programs in improving these students' knowledge and use of the alphabetic system which is the focus of phonics programs. Growth in the reading of miscellaneous words with irregularities was somewhat smaller but significant (d = 0.33), indicating that phonics improved students' ability to read irregularly spelled words, presumably by improving their memory for these words.

In contrast to strong positive effects of phonics instruction on measures of word reading, these programs were not more effective than other forms of instruction in producing growth in spelling (d = 0.09). This effect size was not statistically different from zero. Likewise, phonics programs did not produce significant growth in reading comprehension (d = 0.12) although a small, statistically significant effect was observed on oral reading (d = 0.24).

Because the comparisons involving older children included a large number focusing on disabled readers, the 17 RD comparisons were analyzed separately. Effect sizes proved almost identical to those for the larger group reported in Table 3 (Appendix E) with one important exception. The effect size on the measure of reading comprehension, though small, was statistically greater than zero (d = 0.27, based on eight comparisons that were homogeneous). This indicates that, contrary to the general finding of no effect, systematic phonics instruction did help reading disabled students comprehend text more successfully than nonsystematic/no-phonics programs.

Because most of the comparisons above 1st grade involved poor readers (78%), the conclusions drawn about the effects of phonics instruction on specific reading outcomes pertain mainly to them. Findings indicate that phonics instruction helps poor readers in 2nd through 6th grades improve their word reading skills. However, phonics instruction appears to contribute only weakly, if at all, in helping poor readers apply these skills to read text and to spell words. There were insufficient data to draw any conclusions about the effects of phonics instruction with normally developing readers above 1st grade.

The absence of effects on spelling is noteworthy since the same finding was detected in the Panel's meta-analysis of phonemic awareness instruction. In the PA review, the Panel found that younger readers experienced growth in spelling as a result of phonemic awareness training, but the older disabled readers did not show improvement over controls. One possible explanation is that poor readers experience special difficulty learning to spell (Bruck, 1993). Remediation of this difficulty may require special instruction targeted at spelling. Another explanation may be that as readers move up in the grades, remembering the spellings of words is less a matter of applying letter-sound correspondences and more a matter of knowing more advanced spelling patterns and morphologically based regularities which is not typically addressed in phonics instruction.

Further research is needed to explore the value of phonics instruction in grades beyond 1st grade. Perhaps phonics instruction could be made stronger by combining it with instruction that helps children learn to read words in other ways, specifically, reading words from memory, reading words by analogy to known words, and reading words using spelling patterns and multisyllabic decoding strategies. Some phonics programs in the database did teach children about spelling patterns and the use of an analogy strategy to read words (see results presented below). Also it may be important for phonics programs to include systematic instruction in reading fluency and automaticity when phonics is taught to older students. A few of the programs in the database included exercises to promote fluency. Very likely, phonics programs that emphasize decoding exclusively and ignore the other processes involved in learning to read will not succeed in making every child a skilled reader.

Separation of Reader Ability Groups at Each Grade Level

To clarify whether and how readers with different reading abilities across the different grades responded to phonics instruction, treatment-control group comparisons were grouped by grade and reading ability. There were 62 comparisons with posttests administered when the program was completed or at the end of the first year of the program, whichever came first. Table 5 (Appendix E) shows how these comparisons were distributed across the grade-by-reader-ability cells.

Six groups were formed for the meta-analysis:

- 1st grade normally achieving readers
- 2nd through 6th grade normally achieving readers
- kindergarten children at risk for reading problems
- 1st grade children at risk
- 2nd through 6th grade low achievers
- disabled readers.

More precise grade and age information is given in Table 2 (Appendix D), which lists characteristics of each treatment-control group comparison.

The outcome measure was the overall effect size averaged across the six specific measures. Effect sizes significantly greater than zero were evident for five of the six groups of readers. From Table 3, it is apparent that phonics instruction contributed to growth in reading in all groups but the 2nd through 6th grade low achiever group. Among the at-risk and normal readers in kindergarten and 1st grades, effect sizes were moderate to high, ranging from $\underline{d} = 0.48$ to $\underline{d} = 0.74$. Effect sizes were smaller for 2nd though 6th grade normal readers ($\underline{d} = 0.27$) and disabled readers ($\underline{d} = 0.32$). These findings extend the analysis above by revealing effect sizes for specific reader ability groups at each grade level. Findings indicate that the strong impact of phonics instruction was evident in normally developing 1st graders as well as at-risk kindergartners and 1st graders.

There was one group for whom phonics instruction failed to exert a statistically significant impact on the students' growth in reading. This occurred in the eight comparisons involving low achievers in 2nd through 6th grades ($\underline{d} = 0.15$). Although smaller, the effect size for low achievers did not differ significantly from the effect size of disabled readers ($\underline{d} = 0.32$).

Alternative explanations for the ineffectiveness of phonics instruction with older poor readers in 2nd through 6th grades can be offered. Their reading difficulties may have arisen from sources other than decoding, such as lack of fluency or poor reading comprehension skills (see other sections of the NRP report for elaboration of these reading processes). The fact that the IQs of some of the children in these studies were below normal points to comprehension difficulties as a possibility. Another explanation may be that these children were not given sufficiently intensive phonics instruction to remediate their difficulties. In Table 4 are listed properties of the treatment-control group comparisons involving low achievers. Inspection of the characteristics of these studies reveals that only one provided tutoring, thought to be the most effective way to teach phonics (but see below), whereas seven involved class instruction. However, there may be too few studies of low achieving readers in the database (only eight) to draw firm conclusions. Further research is needed to explore how best to remediate their reading difficulties.

Effects of Phonics Instruction Lasting 2 to 3 Years

The evidence on older readers above 1st grade reviewed so far provides no information about the effects of phonics instruction on older students who began phonics instruction in kindergarten or 1st grade.

However, there is relevant evidence in the database. For four comparisons, phonics instruction was introduced in kindergarten or 1st grade to at-risk readers and continued beyond 1 year (Blachman et al., (1999); Brown & Felton, 1990; Torgesen et al., 1999). These treatment-control group comparisons are listed in Table 4 (Appendix E). At the end of 2nd grade, after 2 to 3 years of instruction, the mean effect size was d = 0.43. This is substantially higher than the mean effect size observed for older children receiving only 1 year of phonics instruction in grades beyond 1st (d = 0.27). Because there are so few cases contributing effect sizes, the results are mainly suggestive. They suggest that when phonics instruction is taught to children at the outset of learning to read and continued for 2 to 3 years, the children experience significantly greater growth in reading at the end of training than children who receive phonics instruction for only 1 year after 1st grade.

SES

One additional characteristic of children was examined as a moderator variable, their socioeconomic status. Two different levels were represented in the database, low SES and middle SES. Also present were studies where SES was stated to vary and studies where it was not given. Table 3 shows that effect sizes were greater than zero in all cases. Phonics instruction exerted its strongest impact on low SES children (d = 0.66). Its impact was somewhat less in middle SES students (d = 0.44) although these two values did not differ statistically. These findings indicate that phonics instruction contributes to growth in reading in both low- and middle-class students.

Characteristics of Phonics Instruction

The treatment-control group comparisons were categorized by the type of systematic phonics instruction taught. In all studies, the programs were identified in sufficient detail to determine that systematic phonics was taught. However, some reports provided less description than others. For programs that were well known or were fully described, the Panel was able to make judgments about their characteristics and fit them into categories. Programs that were not described sufficiently were included in the miscellaneous category. (Publications describing programs are referenced in Appendix C.)

Types of Programs

It is important to recognize that the systematic phonics programs in the database varied not just in the way that the Panel categorized them but also in many other potentially important ways. However, the Panel's choice of categories was limited by the information provided in studies. Most authors mentioned whether the program emphasized synthetic phonics or the teaching of blending using larger subunits of words. However, other properties of programs were not consistently mentioned. Some especially important properties, such as the set of letter-sound relations covered were rarely mentioned. The four categories that were employed are listed in Table 2 (Appendix D) along with the specific treatment-control group comparisons in each category. (For the future, the Panel urges researchers to provide full descriptions of programs that are studied. Journal editors also should insist on this.)

Programs that emphasized systematic synthetic phonics were placed in one category. These programs taught students to transform letters into sounds (phonemes) and to blend the sounds to form recognizable words. This was by far the most common type of program, utilized in 39 of the comparisons. Some of the programs were developed by researchers while others were published programs, some widely used in schools, for example, Jolly Phonics, the Lindamood ADD program, the Lippincott program, Open Court, Orton Gillingham, Reading Mastery (also known as Direct Instruction or DISTAR), and Sing Spell Read & Write.

The second category of programs did not emphasize a synthetic approach at the phonemic level. Rather children were taught to analyze and blend larger subunits of words such as onsets, rimes, phonograms, or spelling patterns along with phonemes. Some of these programs were referred to as embedded code programs because grapheme-phoneme relations were taught in the context of words and text. Teaching children to segment and blend words using onsets and rimes taught them about units as small as graphemes and phonemes because onsets (i.e., the initial consonants in words) are very often single phonemes. In some programs, recognizing rimes in words provided the basis for teaching students the strategy of reading new words by analogy to known words sharing the same rimes. Words in texts were built from linguistic patterns. Writing complemented reading in most programs. The programs in this category included Edmark, Hiebert's embedded code program, three Reading Recovery® programs modified to include systematic phonics, and a program derived from the Benchmark Word Identification program.

One of the 11 studies in the Larger Unit category, that by Tunmer and Hoover (1993), produced an atypical effect size, $\underline{d} = 3.71$, which was much larger than the other effects. It should be noted that this study was atypical in that it was more intensive than most others. It involved one-on-one tutoring by highly trained teachers, and it combined phonemic awareness, phonics, and Reading Recovery® instructional strategies. To reduce the influence of this comparison on the overall mean, its effect size was reduced to equal the next largest effect size in the set, $\underline{d} = 1.41$. (This method of adjusting effect sizes to deal with outliers was only applied in analyses that involved a small number of comparisons.)

The third category, referred to as miscellaneous, consisted of phonics programs that did not fit into the synthetic or larger unit categories. In some studies, the descriptions of programs did not state that a synthetic strategy was taught. If the program was not known to teach this decoding strategy, then it was placed in the miscellaneous category. Also, if the scope of instruction was limited and did not constitute a full phonics program (i.e., Haskell et al., 1992; Lovett et al., 1990), it was considered to be miscellaneous. This set included a spelling program, traditional phonics basal programs, and some researcher-devised instruction that focused on word analysis procedures.

The fourth category, referred to as combination programs, included only two comparisons. However, these could not be fit into the other categories because they examined the effects of teaching two of the other categories, a synthetic phonics program and a larger-units word analogy program (Lovett et al., in press). The comparisons differed in the order that the two programs were taught. The mean effect size for the combined programs was $\underline{d} = 0.42$.

Effect sizes reported in Table 3 show that programs in all three categories produced effect sizes that were significantly greater than zero. This verifies that the three types of phonics programs were more effective than control programs in helping children learn to read. The 39 synthetic phonics programs produced a moderate impact on growth in reading ($\underline{d} = 0.45$). The 11 programs that emphasized larger units created a somewhat smaller impact ($\underline{d} = 0.34$) and likewise the ten miscellaneous programs' effect was smaller ($\underline{d} = 0.27$). However, the three effect sizes did not differ statistically from each other ($\underline{p} > 0.05$). There were relatively few comparisons in the larger unit group. Additional research would be useful for determining whether the small difference between the synthetic and larger unit approaches is a reliable one.

Specific Phonics Programs

There were seven phonics programs that were studied in three or more treatment-control comparisons. The identities of programs and properties of the comparisons testing their effectiveness are listed in Table 6 (Appendix F). Descriptions of the programs are provided in Table 7 (Appendix E). Effect sizes of these comparisons were subjected to a meta-analysis. Results in Table 3 (Appendix F) reveal that all effect sizes were statistically greater than zero, indicating that all the phonics programs produced significantly greater growth in reading than control group programs. The sets of effect sizes for all but one of the programs proved to be homogeneous. Effect sizes ranged from a high of $\underline{d} = 0.68$ for the Lippincott program to a low of $\underline{d} = 0.23$ for the Orton-Gillingham-based programs. Possible reasons for lower effect sizes in the case of Orton Gillingham comparisons are evident in Table 6 (Appendix F). Class-based instruction predominated, and this instruction was tested exclusively with older students (2nd through 6th graders) many of whom were poor readers. These conditions may have made it harder to produce substantial growth in reading.

Although there appear to be sizeable differences in effect sizes distinguishing the programs, the statistical test was not significant. However, drawing the conclusion that these programs are equally effective is premature because there were too few comparisons assessing each program to yield reliable results. Rather, findings should be considered suggestive in need of more studies for verification.

Evaluation of these separate programs was undertaken in the meta-analysis solely because of their prevalence in the database. The programs are older and hence more frequently studied than newer programs. But this does not mean that they are considered to be any better than newer programs that were not analyzed.

Impact of Synthetic Phonics Programs on Different Groups of Readers

Because there were so many comparisons (39) assessing the effects of synthetic phonics programs, it was possible to examine whether this type of program was more beneficial for some grade and reader ability groups than for others. Two groups, at-risk kindergartners and at-risk 1st graders, had the same effect size so they were combined into one group comprising nine comparisons. As evident in Table 3, all groups but one showed effect sizes significantly greater than zero, and all but one group had homogeneous sets of effects. This indicates that synthetic phonics programs produced stronger growth in reading than control programs in most of the different reader groups. Possible reasons why low-achieving readers in 2nd through 6th grades did not benefit were suggested earlier.

Effect sizes varied across the groups. A test to determine whether some groups benefited more from synthetic phonics than other groups showed that effects were significantly greater for at-risk kindergartners and first graders (d = 0.65) than for the two groups of older 2nd through 6th grade readers. These findings indicates that synthetic phonics programs were especially effective for younger, at-risk readers.

Instructional Delivery Unit

Another property of systematic phonics instruction expected to influence growth in reading was the delivery unit. Three types were distinguished. There were eight treatments in which students received one-to-one tutoring. This was expected to be the most effective form of phonics instruction, particularly for low achieving and disabled readers, because it was tailored to individual students. Small group instruction was also expected to be especially effective because attention to individual students was still possible, and in addition, the social setting was expected to enhance motivation to perform and opportunities for observational learning. In the Panel's review of phonemic awareness training studies, findings indicated that effect sizes were significantly greater with small groups than with classrooms or tutoring. Because classrooms involve a much higher ratio of students to teachers, phonics instruction delivered in this setting was expected to be less effective than in the other two settings.

In categorizing studies, it was easiest to determine when tutoring was used because this was clearly stated and described. Identifying whether studies used small groups was also straightforward because training procedures included this descriptive although it was not always clear that this was the only way that instruction was delivered. However, in the case of whole class instructional, sometimes this category was attributed to studies by default. In many reports, descriptions made clear that the phonics program was taught by teachers to their classrooms of students, but the unit of instruction they used to teach the phonics part of programs was not explicitly stated; so, it was inferred to be the class.

Before the meta-analysis was conducted, an adjustment was made to one effect size in the tutoring comparisons. This was considered important because there were only eight comparisons in this set. One of the tutoring studies (Tunmer & Hoover, 1993) produced an atypical effect size, d = 3.71, which was much larger than the other effects. To limit the influence of this comparison on the overall mean, its value was reduced to equal the next largest effect size in the set, d = 1.99.

Results of the analysis of effect sizes for the three types of instructional units revealed that all produced positive effects that were statistically greater than zero, indicating that tutoring, small groups and classes were all effective ways to deliver phonics instruction to students (see Table 3). In addition, the set of effect sizes for comparisons involving small groups was homogeneous, indicating that small group effects are not explained by additional moderator variables and that the mean is a good estimate of the actual effect size, d = 0.43.

Tutoring produced an effect size of d = 0.57 which was greater than the effect size for small groups, d = 0.43, and for classrooms, d = 0.39. However, none of these effects differed statistically from each other. This evidence falls short in supporting the expectation that tutoring would prove especially effective for teaching phonics. However, perhaps there were too few comparisons assessing the effects of tutoring (only eight) to yield reliable findings. On the other hand, it might be noted that the instructional delivery given to the control groups against which tutoring was compared did not involve tutoring in the majority (62%) of the cases. This inequality should have given tutoring an extra advantage. However, it did not.

Inspection of effect sizes for individual studies in Table 2 reveals that some whole class programs produced effect sizes as large, and sometimes larger, than those produced by small groups or tutoring. Given the enormous expense and impracticality of delivering instruction in small groups or individually—except for children who have serious reading difficulties—research is needed to determine what makes whole class phonics instruction effective.

It is interesting to note that the same comparison of instructional units was conducted in the meta-analysis of phonemic awareness training effects. Results showed that small groups were significantly more effective than tutoring or classrooms. Why small groups were more effective for teaching phonemic awareness but not phonics is not clear and awaits further research.

Type of Control Group

To test whether systematic phonics programs produced superior growth in reading, researchers utilized control groups that received unsystematic phonics or no-phonics instruction. The types of control groups chosen by researchers varied across the studies. As mentioned earlier, some studies included more than one type of control group. Selected for analysis were the control groups that were taught the least amount of phonics. These were categorized into five types based on descriptions and labels provided in the studies: basal, regular curriculum, whole language, whole word, and miscellaneous.

Usually basal programs were those already in use at schools. "Regular curriculum" was the label covering cases in which controls received the traditional curriculum or the regular class curriculum in use at schools with no further specification of its contents other that asserting it did not teach phonics systematically. This category covered cases where performance in that grade at that school during previous years was used as a baseline without additional description of the actual program taught. In comparisons involving students identified as at risk by schools, control groups received the standard intervention offered by the schools to treat reading problems.

Whole language was the label used by authors to characterize programs. In two studies (Freppon, 1991; Klesius et al., 1991), the purpose was to examine the effectiveness of whole language programs, not phonics programs that were taught to control groups. In both cases, phonics was taught with a "skill and drill" basal program that was not well described. Control groups that were taught with a Big Books program and with language experience were labeled as whole language.

There were a few programs given to control groups that taught whole words or sight words without much attention to letter-sound relations. These were classed as whole word programs.

Control group programs that did not fit into one of these categories were placed in a miscellaneous category. These included programs teaching traditional spelling, academic study skills, and tutoring in academic subjects. In one case, as a control for parents teaching their own children systematic phonics, the children spent time reading books to their parents (Leach & Siddall, 1990).

Of interest was whether phonics instruction would produce superior growth in reading regardless of the type of control group, and whether phonics instruction would appear more effective when compared to some types of control groups than to others. There were no a priori reasons for expecting effect sizes to be influenced by the type of control group, particularly since the criteria of standard-classroom instruction with minimal phonics had been applied consistently across studies in selecting control groups.

Results in Table 3 (Appendix E) reveal that all of the control groups yielded effect sizes that were statistically greater than zero and all favored the phonics treatment. Effects sizes ranged from $\underline{d} = 0.31$ for whole language controls to $\underline{d} = 0.51$ for whole word controls. Effect sizes for basal and miscellaneous control groups were homogeneous. Additional tests revealed that none of the effect sizes differed significantly from the others. These findings indicate that systematic phonics instruction proved effective regardless of the type of control group that was used.

Design of Studies

Studies in the database varied in methodological rigor. It is important to rule out the possibility that the positive effects of phonics instruction detected in the meta-analysis arose from poorly designed studies. Three features of the studies were coded and analyzed to determine whether more rigorous designs yielded larger or smaller effect sizes: assignment of participants to treatment and control groups, potential presence of pre-experimental differences between groups, and sample size.

Random Assignment

Experimental designs that randomly assign students to treatment and control groups have stronger internal validity than designs that assign already existing groups to the treatment and control conditions. The latter procedure is referred to as nonequivalent group assignment. The goal of experiments is to provide solid evidence that the treatment or lack of it, rather than anything else, explains gains observed in performance following the treatment. Random assignment serves to reduce the likelihood that pre-experimental differences, rather than treatment effects, explain differences between treatment and control groups on outcome measures. When nonequivalent groups are used, statistical techniques can be applied to eliminate pretest differences between groups when outcome measures are analyzed. However, this is not as satisfactory a solution as random assignment.

Most studies in the database provided information regarding how students were assigned to treatment and control groups. If this was not mentioned, then the study was considered to have used nonequivalent groups. Table 3 (Appendix E) shows that studies using random assignment and studies using nonequivalent groups yielded very similar effect sizes, both of which were statistically greater than zero. These findings confirm that the positive effects of systematic phonics instruction did not arise primarily from studies with weaker nonequivalent group designs.

Pre-Experimental Differences

Studies were also coded for the presence of possible or actual pretest differences between treatment and control groups. Effect sizes for questionable studies were calculated separately from studies that were not questionable in this regard. There were 15 comparisons for which no information about pretests was provided and the groups were not randomly assigned. The mean effect size was $\underline{d} = 0.49$. There were ten studies that reported pretest differences and did not use random assignment. The mean effect size in this case was $\underline{d} = 0.37$. When studies containing potential or actual pretest differences were removed from the dataset, effect sizes changed very little and in fact increased slightly, from $\underline{d} = 0.44$ to $\underline{d} = 0.46$. These findings indicate that pretreatment differences between experimental and control groups did not explain why phonics-trained groups outperformed control groups on outcome measures across studies. It was the phonics instruction itself that very likely produced the greater gains in reading.

Sample Size

Another factor indexing the rigor of studies and the reliability of outcomes is sample size, with results of larger studies producing stronger results than smaller studies. The number of students participating in comparisons included in the database varied from 20 to 320. Sample sizes were used to group the comparisons into quartiles, and effect sizes were calculated for each quartile. From Table 3, it is apparent that effect sizes were very similar across quartiles and were all statistically greater than zero. The largest effect size, $\underline{d} = 0.49$, emerged in studies having the largest samples. These findings show that the positive effects of systematic phonics instruction were not limited to studies that produced effects with relatively few students.

Discussion

Findings of the meta-analysis allow us to conclude that systematic phonics instruction produces gains in reading and spelling not only in the early grades (kindergarten and 1st grades) but also in the later grades (2nd through 6th grades) and among children having difficulty learning to read. Effect sizes in the early grades were significantly larger (d = 0.55) than effect sizes above 1st grade (d = 0.27). These results support Chall's (1967) assertion that early instruction in systematic phonics is especially beneficial to growth in reading.

Although there was some thought that kindergartners might not be ready for phonics instruction because they first need to acquire extensive knowledge about how print works (e.g., Stahl & Miller, 1989; Chall, 1996a, b), findings did not support this possibility. Phonics instruction produced similar effect sizes in kindergarten (d = 0.58) and 1st grade (d = 0.54).

Phonics instruction can be described in terms of the method used to teach children about letter-sound relations and how to use letter-sounds to read or spell. There are synthetic, analytic, analogy, spelling-based, and embedded approaches to teaching phonics. Phonics instruction can also be described in terms of the content covered, for example, short vowels, long vowels, digraphs, phonics generalizations, onsets and rimes, phonograms, and so forth. In the present meta-analysis, only the types of methods were compared in terms of the effect sizes produced, and no significant differences among methods were detected.

Stahl et al. (1998) suggest that the benefits of phonics instruction and differences among phonics approaches may arise from the amount of content covered and learned by students rather than from properties distinguishing the various methods. Synthetic methods tend to be efficient in covering content and tend to cover an ambitious number of sound-symbol correspondences in the 1st grade year. Other approaches vary considerably in the amount that they cover. To understand phonics instruction and its effects on student learning, research is needed to study separately the effects of teaching methods from the effects of content coverage. Systematic phonics instruction is focused on teaching children the alphabetic system and explicitly how to apply it to read and spell words. Phonics skills would be expected to show effects on text comprehension to the extent that phonics skills help children read the words in texts. This is one reason why phonics instruction may have exerted less impact on text comprehension outcomes than on word reading outcomes, because the impact is indirect. In addition, although phonics programs do give children practice reading connected text, the purpose of this practice is centered on word recognition rather than on comprehending and thinking about the meaning of what is being read. This may be another reason why effect sizes on text comprehension were smaller than effect sizes on word reading.

In the present analysis, systematic phonics instruction exerted a lower than expected impact on reading growth in low achieving readers (d = 0.15) and disabled readers (d = 0.32). The Panel's meta-analysis of phonemic awareness training studies included comparisons involving poor readers. Most of these studies would qualify as phonics studies because letter-sound manipulations were part of the phonemic awareness training. The studies were not included in the phonics database in order to avoid duplication of studies across meta-analyses. The effect size on reading outcomes in the PA meta-analysis involving poor readers was d = 0.45, a value quite a bit higher than the effect sizes produced by phonics instruction. It may be that including more phonemic awareness training with letters might improve the quality of phonics instruction given to poor readers. However, there may be other factors that explain the difference as well. Closer scrutiny of the two sets of studies is needed to identify possible reasons. For example, RD students in the phonics analysis may have been older than students in the PA analysis.

The overall effect size of systematic phonics instruction in 1st grade was d = 0.54. Although moderate in size, this value is somewhat low when compared to effect sizes found in other similar reviews. Stahl and Miller (1989) conducted a meta-analysis of phonics instruction and drew their comparisons from the Cooperative First Grade Studies (Bond & Dykstra, 1967, 1998) whose participants should be similar to 1st graders in the present database. Stahl and Miller found effect sizes of 0.91 on the Stanford Word Reading subtest and 0.36 on the Paragraph Meaning subtest for children who received phonics instruction similar to that studied here. Overall, these are higher effect sizes than those detected in the present meta-analysis.

The discrepancy may arise from differences in the way the Panel created its database. Whereas the Panel's review was limited to studies published in peer-reviewed journals, authors of the previous meta-analyses made a great effort to find "fugitive" or unpublished studies to include. One reason to search widely for studies is that the publishing process tends to screen out studies reporting null effects, and this runs the risk of biasing the data set towards positive effects. However, such a bias would be expected to favor a larger effect size using National Reading Panel procedures, and this did not happen. Another possible reason for the discrepancy is that the previous analyses included unpublished studies, thus running the risk of admitting studies of poor quality with inflated effect sizes. Limiting studies to those passing the test of peer review minimizes this risk.

Another possible explanation for the Panel's smaller effect size is that the database involved more recent studies. There may have been more of a tendency for later studies to focus on at-risk, low-achieving, and disabled readers for whom growth in reading may be harder to achieve. Perhaps the reading instruction experienced by students in control groups included more phonics than the reading instruction received by control groups in earlier years. In the 1960s, basal readers used a whole word methodology whereas the control conditions in more recent studies are presumably more eclectic. Table 2 identifies the control groups used by studies in the corpus. Whereas some groups were true "no-phonics" controls, other groups received some phonics instruction. It may be that, instead of examining the difference between phonics instruction and no phonics instruction, a substantial number of studies actually compared more systematic phonics instruction to less phonics instruction. This would produce smaller differences between treatment and control groups and hence smaller effect sizes.

In one of the studies in the database, Evans and Carr (1985) conducted extensive observations of the instruction received by treatment and control groups and reported their observations numerically. They found that the phonics classes spent 13.38% of the group time and 11.94% of independent work time on word analysis, whereas the control group spent 5.37% of the group time and 1.84% of the independent time on word analysis. Although there is a difference favoring the phonics group, the finding shows that control classes did spend some time on word analysis as well. Chall and Feldmann (1966) found that there was considerable variation in instruction, even in classes professing to be using the same methods. This underscores the importance of researchers taking steps not only to assess outcomes of instructional treatments but also to document the nature of the instruction received by treatment and control groups to verify whether and how they actually differed.

Studies to Illustrate Systematic Phonics Instruction and Its Contribution to Growth in Reading

Some of the studies in the database are described to provide a glimpse of the experiments contributing effect sizes and to portray various types of phonics instruction that were examined.

Phonics Instruction in Kindergarten

Systematic phonics instruction in kindergarten was studied in six articles. The main goals included teaching children the shapes of letters and their sounds, how to analyze sounds in words (phonemic awareness), and how to use letter sounds to perform various reading or writing tasks appropriate for children just starting out. In the study by Stuart (1999), three kindergarten teachers utilized the Jolly Phonics program (Lloyd, 1993), and three teachers centered their instruction around Holdaway's (1979) Big Book approach. Teachers taught these programs 1 hour per day for 12 weeks during the latter half of kindergarten.

Big Book instruction included work with letters. Teachers drew children's attention to written words in the books and they talked about letters in words. Also, teachers employed various "imaginative and fun activities" to help children learn letters and their sounds. However, the instruction was not systematic; the sequence of teaching letters was not prescribed, and no special system for remembering letter-sound relations was taught.

The Jolly Phonics program was more systematic and prescribed in its teaching of letters. This program was developed by Lloyd (1993), a teacher, for 4- and 5-year-olds in their first year of schooling in the United Kingdom. Central to the program is the use of meaningful stories, pictures, and actions to reinforce recognition and recall of letter-sound relationships, and precise articulation of phonemes. There are five key elements to the program: (1) learning the letter sounds, (2) learning letter formation, (3) blending for reading, (4) identifying the sounds in words for writing, and (5) tricky words that are high frequency and irregularly spelled. The program includes activities and instruction specifically designed to address those skills most needed in the development of early literacy. Unlike many older phonics approaches, however, Jolly Phonics promotes playful, creative, flexible teaching that fits well with whole language practice and leads directly to authentic reading and writing.

At the end of training in either Jolly Phonics or Big Books, children were given various tests to compare effects of the programs. Results showed that Jolly Phonics at-risk kindergartners were able to read significantly more words and pseudowords and to write more words than the Big Book group. The overall effect size was $d = 0.73$. A year later, the children were retested. The Jolly Phonics group outperformed the control group in reading and spelling words but not in reading comprehension. These results show that phonics instruction in kindergarten is effective in boosting children's progress in learning to read and write words.

One interesting feature of the Jolly Phonics program is that children are taught hand gestures to help them remember the letter-sound associations. For example, they make their fingers crawl up their arm portraying an ant as they chant the initial sound of "ant" associated with the letter a. The value of mnemonics for teaching letter-sound relations to kindergartners is supported by evidence. In a study by Ehri, Deffner, and Wilce (1984), children were shown letters drawn to assume the shape of a familiar object, for example, s drawn as a snake, h drawn as a house (with a chimney). Memory for the letter-sound relations was mediated by the name of the object. Children were taught to look at the letter, be reminded of the object, say its name, and isolate the first sound of the name to identify the sound (i.e., s - snake - /s/). With practice they were able to look at the letters and promptly say their sounds. Children who were taught letters in this way learned them better than children who were taught letters by rehearsing the relations with pictures unrelated to the letter shapes (e.g., house drawn with a flat roof and no chimney) and also better than children who simply rehearsed the associations without any pictures.

Application of this principle can be found in Letterland (Wendon, 1992), a program that teaches kindergartners letter-sound associations. In this program, all the letters are animate characters that assume the shape of the letters and have names prompting the relevant sound, for example, Sammy Snake, Hairy Hat Man, Fireman Fred, Annie Apple. The task of learning the shapes and sounds of all the alphabet letters is difficult and time-consuming, particularly for children who come to school knowing none. The relations are arbitrary and meaningless. Techniques to speed up the learning process are valuable in helping kindergartners prepare for formal reading instruction.

The motivational value of associating letters with interesting characters or hand motions and incorporating this into activities and games that are fun is important for promoting young children's learning. If the task of teaching letters is stripped bare to one of memorizing letter shapes and sounds, children will become bored and easily distracted and will take much longer to learn the associations.

A Developmental Approach to Phonics Instruction in Kindergarten

Another phonics program for kindergartners was studied by Vandervelden and Siegel (1997). The interesting feature of their approach was to tailor the intervention to individual children's level of knowledge. This is important because kindergartners vary greatly in how much they already know about letters when they enter school. The instruction lasted 12 weeks, with children receiving two sessions per week. There were 15 children that received phonics instruction and 15 that received the same instructional format but focused on classroom activities and materials. Children were pretested. The three children who showed the least knowledge received one-on-one tutoring, the next eight lowest scoring children were instructed in pairs, and the four highest scoring children worked in a small group.

The skills taught to phonics-treated children who lacked them included the following: learning sounds for consonant letters; use of initial letter-sound matches to recognize, spell, and read words; segmenting words into sounds and spelling the sounds; orally reading text containing the words learned in this way; learning correct spellings of words by analyzing letter-sound constituents; and use of rime analogy in reading and spelling words. Easier skills were taught before harder skills. Instruction began at levels appropriate for individual learners.

In the control group, children engaged in activities used in their classrooms. This included letter learning and phonemic awareness. However, children were not explicitly guided in the use of these skills to read and write.

Results showed that the phonics groups outperformed the control group on tests of phonemic awareness and letter-sound relations but not letter names. Also, the phonics group did better on tests of speech-print matching of words and pseudowords (e.g., which written word, milk, monk, or mask says "mask"), on tests of writing the sounds in words, and on some but not all measures of word reading. The overall effect size was $d = 0.47$. It is important to recognize, however, that these kindergartners were still at a rudimentary level in their development as readers. For example, at the end of the treatment, they were able to match 43% of the written and spoken words correctly; they read only a mean of 10 out of 60 high frequency words such as up, yes, and book, and they spelled only 46% of the sounds in words. This suggests that teaching students to use phonics skills to read and spell words at the kindergarten level may yield only limited success. However, perhaps this program was not optimally designed or did not last long enough.

A 2.5-Year Phonics Program Beginning With Phonemic Awareness

A lengthier, more comprehensive program lasting more than 2 years was studied by Blachman et al. (1999). Classroom teachers used the program with low SES, inner-city children. Instruction began in kindergarten with a focus on phonemic awareness training lasting 11 weeks. In 1st grade, explicit, systematic instruction in the alphabetic code was taught. This instruction continued in 2nd grade for children who did not complete the program in 1st grade. Control children participated in the school's regular basal reading program that included a phonics workbook that children used independently.

The phonemic awareness instruction taught children to perform a "say it and move it" procedure in which they moved a disk down a page as they pronounced each phoneme in a word. They practiced segmenting two- and three-phoneme words in this way. Then a limited set of eight letter-sound relations was taught, and children moved the letters rather than the disks. It is noteworthy that when children began this program, they knew on average only two letter sounds and could not yet write their names. Thus, the participants were starting from zero in their alphabetic learning. By the end of kindergarten, children knew on average 19 letter names and 13 letter-sounds, indicating that substantial learning had occurred.

At the beginning of 1st grade, there was still wide variation in children's letter knowledge and phonemic awareness. This underscores the fact that even though children receive the same instruction, they still differ in how quickly they learn what they are taught. To address the variation, children were assigned to ability groups. The core of the reading program involved daily, 30-minute lessons consisting of five steps that emphasized the alphabetic code:

1. Teaching new sound-symbol correspondences with vowels highlighted in red
2. Teaching phoneme analysis and blending
3. Reading regularly spelled, irregularly spelled, and high-frequency words on flash cards to develop automaticity
4. Reading text containing phonetically controlled words
5. Writing four to six words and a sentence to dictation.

By the end of the program, children had been introduced to all six syllable types: closed (fat), final E (cake), open (me), vowel team (pain), vowel + r (burn), and consonant le (table). Vocabulary development and work on reading comprehension was incorporated as well, with more time spent reading text as the year progressed and children's reading vocabulary grew.

Inservice workshops held once a month were used to instruct teachers how to implement the program. The instruction presented information about how children acquire literacy skills and the role of phonological processes in learning to read. Teachers learned how to provide explicit instruction in the alphabetic code. The issue of pacing was stressed. Developing students' phonemic awareness, letter-sound knowledge, and word recognition skills was identified as being more important than "covering the material."

To assess how far children had progressed in their reading and writing, various tests were given at the end of kindergarten, 1st grade, and 2nd grade. Results showed that kindergartners receiving PA training outperformed control students, with d = 0.72. At the end of 1st grade, children who received explicit phonics training achieved significantly higher scores than controls, with d = 0.64. During 2nd grade, children in the phonics group who had not met the program's goals received additional instruction while the rest received regular classroom instruction. On posttests at the end of the year, the phonics-trained group continued to outperform the control group, with d = 0.36.

These findings show that the explicit systematic instruction in phonics provided by the Blachman program improved low SES children's ability to read words more than a basal program less focused on teaching children alphabetic knowledge and word reading skills. Several features of this program are noteworthy and may underlie its effectiveness. The same program continued over three grades, thus insuring consistency and continuity in children's learning the alphabetic system and how to use it to read and spell. The program began in kindergarten with alphabetic code instruction that was appropriate for children's level of knowledge. They were taught phonemic awareness and a limited set of letter-sound relations which they used to make and break words. Both PA and letter knowledge are known to be the strongest predictors of how well children will succeed in learning to read. Delivery of instruction was tailored to enable all students to complete the program. Tests were given to assess children's progress and to distinguish those children who needed further instruction from those who did not. Instruction in the alphabetic code included various kinds of reading and writing skills, not only sounding out and blending words but also building memory for words, spelling words, and reading words in text. An extensive set of letter-sound relations including vowels was taught and applied to various types of words organized by syllable structure. Teachers were provided with inservice workshops during the school year to help them not only provide instruction correctly but also to understand the reading processes and their course of acquisition in students. These properties of the Blachman phonics program may account for its effectiveness. Further research to examine the contribution of such properties is needed.

An Intensive 3-Year Tutoring Program: Synthetic vs. Embedded Phonics Instruction

Another study in the database, by Torgesen et al. (1999), also provided phonics instruction throughout the primary grades. In this study, two different forms of phonics instruction were compared, one which provided very explicit and intensive instruction in PA and phonemic decoding called PASP (phonological awareness plus synthetic phonics), while the other provided systematic but less explicit instruction in phonemic decoding in the context of more instruction and practice in text comprehension, called EP (embedded phonics). Instruction was provided by tutors rather than classroom teachers. Kindergarten children with poor PA and letter knowledge received 88 hours of tutoring over 2.5 years, with sessions lasting 20 minutes and scheduled four times per week. Instruction was individually paced according to the progress that children made. This instruction was added to the reading instruction they received in the classroom. There were two control groups, one that received tutoring that supported regular classroom instruction, and one in which children received only regular classroom instruction. Instruction in the tutoring control condition included some phonics oriented activities. There were 180 children from 13 schools. Children were randomly assigned to one of the four conditions.

The PASP children received the Auditory Discrimination in Depth program (Lindamood & Lindamood, 1984). This program began by teaching children phonemic awareness in a unique way. Children were led to discover and label the articulatory gestures associated with each phoneme by analyzing their own mouth movements as they produced speech. For example, children learned that the word beat consists of a lip popper, a smile sound, and a tongue tapper. Children learned to track the sounds in words with mouth pictures as well as colored blocks and letters. Most of the time in this program was spent building children's PA and their decoding skills although some attention was given to the recognition of high frequency words, text reading, and comprehension.

The EP program began by teaching children to recognize whole words. Instruction in letter-sounds occurred in the context of learning to read words from memory (by sight). Also, children wrote sentences and read what they wrote. In this context, phonemic awareness was taught by having children segment the sounds in words before writing them. When children had sufficient reading vocabulary, they began reading short stories to build their reading vocabulary further. The emphasis was on acquiring word level reading skills, including sight words and phonemic decoding skills. Also, attention was given to constructing the meanings of stories that were read.

One step taken in the Torgesen et al. study was to videotape 25% of the PASP and EP tutorial sessions and analyze the interaction to verify how phonics instruction differed in the two programs. The percentages of time spent on the following types of activity were

- PA, letter-sounds, phonemic reading/writing of words: 74% (PASP) vs. 26% (EP)
- Sight word instruction: 6% (PASP) vs. 17% (EP)
- Reading/writing connected text: 20% (PASP) vs. 57% (EP).

In comparing the groups' performance on outcomes measures across the grades, Torgesen et al. found that the PASP group read significantly more real words and nonwords and spelled more words than one or both of the control groups. However, the EP group did not outperform the control groups on any of the measures. There was a significant overall effect of interventions on the comprehension measures, but individual contrasts between groups were not statistically significant. Comparison of the PASP and EP groups revealed superior performance by PASP on measures of phonological awareness, phonemic decoding accuracy and efficiency, and word reading accuracy. However, the groups did not differ in word reading efficiency (taking account of speed as well as accuracy) or in the individual contrasts for reading comprehension. Thus, findings revealed that intensive training in phonics produced superior word reading skills compared to embedded phonics training or training given to control groups. Interestingly, neither of the two instructional control groups, embedded phonics or supported classroom instruction, produced significant effects compared to the no treatment control group, while the explicit PASP group did. Based on comparisons to the classroom control group, effect sizes for the two phonics groups were

- PASP: $\underline{d} = 0.33$ (kindergarten), 0.75 (1st grade), 0.67 (2nd grade)
- EP: $\underline{d} = 0.32$ (kindergarten), 0.28 (1st grade), 0.17 (2nd grade).

Clearly, effects of synthetic phonics instruction persisted more strongly over the grades than effects of embedded phonics instruction. Left unclear is whether PASP's effectiveness resulted from the greater time spent teaching alphabetic and phonological processes, or the specific content of the instruction, or a combination of both factors.

Although the comparisons between individual groups were not significant for the comprehension measures, when the outcomes for the PASP group were compared to those of the EP and RCS groups combined, the effect size for the passage comprehension test of the Woodcock Reading Mastery Test-Revised was 0.43. The corresponding effect size for the comprehension measure for the Gray Oral Reading Test–3 was 0.21. While reading comprehension depends upon other processes besides word reading, one would expect to see transfer, particularly in the primary grades where text reading is heavily influenced by word recognition skills. One possible explanation is that the tests of comprehension were standardized and hence were not sufficiently sensitive to detect small within-grade differences. This is because standardized tests are designed to detect differences across the whole range of grades; so, there are only a small number of items at each grade level. Another possibility is that compensatory processes are sufficiently strong to dilute the contribution that superior word recognition skill makes to text reading. That is, children read and comprehend text by utilizing their linguistic and background knowledge combined with their word reading skill. When word reading skill is somewhat weaker, children can rely more heavily on their knowledge about the subject and memory for what they have read to still make sense of the text.

The kindergartners selected to be tutored in reading in the Torgesen et al. (1999) study were severely at risk for becoming disabled readers based on very poor letter knowledge and phonemic awareness which are the two best kindergarten-entry predictors of future reading achievement (Share et al., 1984). However, these children varied greatly in verbal intelligence, with IQs ranging from 76 to 126 in kindergarten and from 57 to 130 in 2nd grade. Thus, the sample in this study included two kinds of potentially poor readers, children who were *unexpectedly* poor readers because their IQs were *higher* than their reading potential scores and children whose below-average reading was not surprising given that their IQs were also below average. These two types of poor readers have been distinguished in other studies by researchers. Various labels, such as dyslexic or learning disabled or reading disabled, have been applied to children whose higher IQs are discrepant with their lower reading skill. Children whose lower reading scores are consistent with their lower IQs have been called low achievers or garden variety poor readers. These children would be expected to display low achievement not only in reading but also in other academic areas requiring cognitive or verbal capabilities.

Torgesen et al. (1999) observed that children in their study varied greatly in their response to instruction. Even in the strongest phonics group, almost one-fourth of the children remained significantly impaired in their decoding or word reading ability at the end of instruction. Torgesen et al. conducted a regression analysis to examine what characteristics of the children predicted how well or poorly they responded to instruction as indexed by their growth in word reading over 2.5 years. They found that the important variables explaining growth were home background (parent occupation and education), kindergarten classroom behavior (activity level, attention, adaptability, social behavior) and phonological capabilities (i.e., phonemic awareness, short-term memory, naming speed). The variable involving IQ differences among the children did not explain any further growth over and above these other variables. Torgesen et al. suggest that whether or not children's IQ is discrepant with their reading potential is probably not relevant in determining their need for special help in acquiring word reading skills.

Modified Reading Recovery® Studies

There were three studies in the database that adopted the Reading Recovery® (RR) format developed by Clay (1993) and altered it to include more systematic work in phonics. The type of phonics instruction involved an emphasis on larger subunits as well as phonemes. The RR program developed by Clay is adminstered by a tutor to children who have fallen behind in reading after a year of instruction. The 30-minute RR lesson includes several activities: rereading two familiar books, reading the previous day's new book, practicing letter identification, writing a story by analyzing sounds in words, re-assembling the words of a cut-up story, reading a new book.

Greaney, Tunmer, and Chapman (1997) modified the RR program by providing explicit instruction in letter-phoneme patterns once children had learned the majority of letters. This work consumed 5 minutes of each session and was substituted for the letter segment of the RR lesson. Children were taught to read pairs of nouns containing common spellings of rimes (e.g., m-eat) and then words with the rime embedded in it (e.g., h-eat-er). They practiced reading and also writing words with these larger rime units referred to as "eggs" because the unit was written in an egg-shaped space. Attention was drawn to the egg units and their utility for reading words. During the final book reading segment of each session, children were encouraged to use the eggs to identify unfamiliar words in the book. This treatment was referred to as rime analogy training. Children in the control group followed the same RR format and read the same words. However, no attention was drawn to rime units in the words, and the words were mixed up rather than taught in sets having the same rimes.

The study was conducted in New Zealand. Both the modified RR and the unmodified RR programs lasted for 12 weeks. The children in the study were from grades 2 through 5 and were the poorest readers in their class. Results showed that the children who received rime training outperformed control children on tests of word and pseudoword reading but not on tests of reading comprehension. The overall effect size was $\underline{d} = 0.37$. These findings reveal that the rime-analogy phonics program produced greater growth in word reading than the whole word program.

Tunmer and Hoover (1993) performed a similar study in which the letter segment of the RR lesson was replaced by more systematic phonics instruction. Children were taught to make, break, and build new words that had similar letters and sounds. Instruction began by focusing on phonograms or rime spellings in words (e.g., make, bake, cake, take). A metacognitive strategy training approach rather than a skill and drill approach was used to make children aware of how letters and sounds work in words and how to use their alphabetic knowledge to read and spell.

Two control groups were included in the study. One group received unmodified RR lessons. The other group received the standard treatment given to poor readers by the school district. This was a pull-out program in which teachers worked with children in small groups. Some word analysis activities were included. The children were all 1st graders in their 2nd year of reading instruction. They were the poorest readers in their class. Posttests were given when RR children achieved the goals of the program. Results showed that the modified RR group outperformed the group receiving the standard small-group instruction on all measures. The overall effect size was $d = 3.71$, indicating that the modified RR phonics program produced an enormous advantage over the treatment received by the standard control group.

In contrast, the modified RR group performed very similarly to the unmodified RR group on the reading measures following training. The only difference was that it took significantly fewer sessions for the modified RR group to achieve the goals of RR than the unmodified RR group. The effect size showing the advantage in reduced time was $d = 1.40$. The same advantage in time, but not in reading outcomes, was uncovered by Iversen and Tunmer (1993) who conducted a very similar study. (The Iversen and Tunmer data were included in the Panel's meta-analysis of phonemic awareness instruction.) Findings of both studies show that Clay's Reading Recovery® program produced the same growth in reading even though it provided less systematic phonics instruction than the modified program and provided it mainly through writing exercises rather than decoding activities. Although reading outcomes were the same, the fact that one program took less time makes the more intensive phonics approach preferable. Because the RR program requires one-on-one tutoring delivered in schools by a few highly trained RR teachers, it is expensive; so, a savings in time can mean either that more students are helped or that fewer teachers are required.

A third study in the database also modified the RR format to include more systematic phonics instruction. In the study by Santa and Hoien (1999), at-risk 1st graders received tutoring that involved story reading, writing, and phonological skills based on a program developed by Morris (1992). The unique part of this phonics program was that it used word study activities to develop phonological awareness and decoding skill. Word study consumed 5 to 6 minutes of the 30-minute lesson. Children were given cards to sort into categories. They might sort picture cards that shared the same initial sounds, or word cards sharing the same vowel sounds. The typical sort involved three patterns with four words in each pattern. Initially, children worked with phonograms (e.g., -at in hat, cat, sat, rat) and then advanced to shared phonemes as the basis for sorting words. Children also were taught to spell by writing letters for the sounds heard in words. Metacognitive strategies were taught including an analogy strategy in which children were urged to use words they know to read words they don't know.

The control group received small group, guided reading instruction. They practiced reading and rereading books in 30-minute lessons but did not receive any word study activities. Results showed that the word study program produced much greater growth in reading than the guided reading program, $d = 0.76$. Gains were greater in reading comprehension as well as word reading. These findings provide evidence for the effectiveness of teaching children phonics through the use of larger units along with phonemes.

Systematic Phonics to Remediate the Reading Difficulties of Disabled Readers

Children who have been diagnosed as reading disabled have severe reading difficulties that are not explained by low intelligence. Systematic phonics programs have been developed to remediate their reading difficulties. RD children have special problems in acquiring word reading skills. Not only do they struggle to read pseudowords, but they also have trouble remembering how to read words they have read before.

Maureen Lovett and her associates (Lovett et al., 1994; Lovett & Steinbach, 1997; Lovett et al., in press) have conducted several studies to examine how to improve the word reading skills of severely disabled readers. They have explored the effectiveness of two types of phonics programs, a synthetic program they call PHAB and a larger-unit program, which teaches children to use subparts of words they know to read new words, referred to as WIST.

The PHAB synthetic phonics program adopted the Direct Instruction model developed by Engelmann and his colleagues (see Appendix) to remediate the decoding and phonemic awareness difficulties of the disabled readers. Children were taught to segment and blend words orally. They were taught letter-sound associations in the context of word recognition and decoding instruction. The program taught a left-to-right decoding strategy to sound out and blend letters into words. Special marks on letters and words provided visual cues to aid in decoding, such as symbols over long vowels, letter size variations, and connected letters to identify digraphs. Cumulative, systematic review and many opportunities for overlearning were used. New material was not introduced until the child had fully mastered previously instructed material. Children were taught in small groups.

The larger-unit, word analogy program called WIST was adapted from the Benchmark Word Identification/ Vocabulary Development program developed by Gaskins et al. (1986). This program had a strongly metacognitive focus. It taught children how to use four metacognitive strategies to decode words: reading words by analogy, detecting parts of words that are known, varying the pronunciations of vowels to maintain flexibility in decoding attempts, and "peeling off" prefixes and suffixes in words. Children learned a set of 120 key words exemplifying high-frequency spelling patterns, five words per day. They learned to segment the words into subunits so that they could use known words and their parts to read other similarly spelled words. They learned letter-sound associations for vowels and affixes. Various types of texts provided children with practice applying the strategies that were taught.

The children participating in the studies were referred to Lovett's clinic because they had severe reading problems. Children were randomly assigned to receive the PHAB program, the WIST program, or a non-reading control program that involved teaching students academic survival skills such as organization and problem solving relevant to the classroom. The students ranged in age from 6 to 13 years or grades 2nd through 6th. The three programs took the same amount of time. In one study, it was 35 hours; in another study, 70 hours.

To evaluate the effectiveness of the programs, performance of students receiving either PHAB or WIST were compared to performance of the control group. There were four comparisons assessing effects of PHAB and four assessing WIST in the database. Although the effect sizes were somewhat variable, the average effect size across the comparisons indicated that both programs produced about the same growth in reading, $\underline{d} = 0.41$ for PHAB and $\underline{d} = 0.48$ for WIST. In two of the comparisons, both reading comprehension and word reading were measured. Substantial gains were evident on both measures. These findings indicate that the two approaches to teaching systematic phonics, one teaching synthetic phonics, and one teaching the use of larger subunits of words to read by analogy, were quite effective in helping disabled readers improve their reading skills.

Conclusions

There were 38 studies from which 66 treatment-control group comparisons were derived. Although each comparison could contribute up to six effect sizes, one per outcome measure, few studies did. The majority (76%) of the effect sizes involved reading or spelling single words, whereas 24% involved text reading. The imbalance favoring single words is not surprising given that the focus of phonics instruction is on improving children's ability to read and spell words. Studies limiting instructional attention to children with reading problems accounted for 65% of the comparisons, 38% involving poor readers considered "at risk" or low achieving and 27% diagnosed as reading disabled (RD). Studies involving 1st graders were overrepresented in the database, accounting for 38% of the comparisons. Fewer kindergartners (12%) and children in 2nd through 6th grades (23%) were represented. Children in the RD group spanned several ages and grades, ranging from ages 6 to 13 and grades 2nd through 6th. Most of the studies (72%) were recently conducted, in the past 10 years.

Systematic phonics instruction typically involves explicitly teaching students a prespecified set of letter-sound relations and having students read text that provides practice using these relations to decode words. Instruction lacking an emphasis on phonics instruction does not teach letter-sound relations systematically and selects text for children according to other principles. The latter form of instruction includes whole-word programs, whole language programs, and some basal reader programs.

The meta-analyses were conducted to answer several questions about the impact of systematic phonics instruction on growth in reading when compared with instruction that does not emphasize phonics. Findings provided strong evidence substantiating the impact of systematic phonics instruction on learning to read.

1. Does systematic phonics instruction help children learn to read more effectively than unsystematic phonics instruction or instruction teaching no phonics?

Children's reading was measured at the end of training if it lasted less than a year or at the end of the first school year of instruction. The mean overall effect size produced by phonics instruction was significant and moderate in size ($\underline{d} = 0.44$). Findings provided solid support for the conclusion that systematic phonics instruction makes a more significant contribution to children's growth in reading than do alternative programs providing unsystematic or no phonics instruction.

2. Are some types of phonics instruction more effective than others? Are some specific phonics programs more effective than others?

Three types of phonics programs were compared in the analysis: (1) synthetic phonics programs that emphasized teaching students to convert letters (graphemes) into sounds (phonemes) and then to blend the sounds to form recognizable words; (2) larger-unit phonics programs that emphasized the analysis and blending of larger subparts of words (i.e., onsets, rimes, phonograms, spelling patterns) as well as phonemes; and (3) miscellaneous phonics programs that taught phonics systematically but did this in other ways not covered by the synthetic or larger-unit categories or were unclear about the nature of the approach. The analysis showed that effect sizes for the three categories of programs were all significantly greater than zero and did not differ statistically from each other. The effect size for synthetic programs was $\underline{d} = 0.45$; for larger-unit programs, $\underline{d} = 0.34$; and for miscellaneous programs, $\underline{d} = 0.27$. The conclusion supported by these findings is that various types of systematic phonics approaches are more effective than non-phonics approaches in promoting substantial growth in reading.

There were seven programs that were examined in three or more treatment-control group comparisons in the database. Analysis of the effect sizes produced by these programs revealed that all were statistically greater than zero and none differed statistically from the others in magnitude. Effect sizes ranged from $\underline{d} = 0.23$ to 0.68. In most cases there were only three or four comparisons contributing effect sizes, so results may be unreliable. The conclusion drawn is that specific systematic phonics programs are all more effective than non-phonics programs and they do not appear to differ significantly from each other in their effectiveness although more evidence is needed to verify the reliability of effect sizes for each program.

3. Is phonics taught more effectively when students are tutored individually, when they are taught in small groups, or when they are taught as classes?

All three delivery systems proved to be effective ways of teaching phonics, with effect sizes of $\underline{d} = 0.57$ (tutoring), $\underline{d} = 0.43$ (small group), and $\underline{d} = 0.39$ (whole class). All effect sizes were statistically greater than zero, and no one differed significantly from the others. This supports the conclusion that systematic phonics instruction is effective when delivered through tutoring, through small groups, and through teaching classes of students.

4. Is phonics instruction more effective when it is introduced to students not yet reading, in kindergarten or 1st grade, than when it is introduced in grades above 1st after students have already begun to read?

Phonics instruction taught early proved much more effective than phonics instruction introduced after 1st grade. Mean effect sizes were kindergarten d = 0.56; 1st grade d = 0.54; and 2nd through 6th grades d = 0.27. The conclusion drawn is that systematic phonics instruction produces the biggest impact on growth in reading when it begins in kindergarten or 1st grade before children have learned to read independently. To be effective, phonics instruction introduced in kindergarten must be appropriately designed for learners and must begin with foundational knowledge involving letters and phonemic awareness.

5. Is phonics instruction beneficial for children who are having difficulty learning to read? Is it effective in preventing reading failure among children who are at risk for developing reading problems in the future? Is it effective in remediating reading difficulties in children who have been diagnosed as reading disabled and children who are low-achieving readers?

Phonics instruction produced substantial reading growth among younger children at risk of developing future reading problems. Effect sizes were d = 0.58 for kindergartners at risk and d = 0.74 for 1st graders at risk. Phonics instruction also improved the reading performance of disabled readers (i.e., children with average IQs but poor reading) for whom the effect size was d = 0.32. These effect sizes were all statistically greater than zero. However, phonics instruction failed to exert a significant impact on the reading performance of low-achieving readers in 2nd through 6th grades (i.e., children with reading difficulties and possibly other cognitive difficulties explaining their low achievement). The effect size was d = 0.15, which was not statistically greater than chance. Possible reasons might be that the phonics instruction provided to low-achieving readers was not sufficiently intense, that their reading difficulties arose from sources not treated by phonics instruction such as poor comprehension, or that there were too few cases (i.e., only eight treatment-control comparisons pulled from three studies) to yield reliable findings.

The conclusion drawn from these findings is that systematic phonics instruction is significantly more effective than non-phonics instruction in helping to prevent reading difficulties among at-risk students and in helping to remediate reading difficulties in disabled readers. No conclusion is drawn in the case of low-achieving readers because it is unclear why systematic phonics instruction produced little growth in their reading and whether the finding is even reliable. Further research is needed to determine what constitutes adequate remedial instruction for low-achieving readers.

6. Does systematic phonics instruction improve children's reading comprehension ability as well as their decoding and word-reading skills?

Systematic phonics instruction was most effective in improving children's ability to decode regularly spelled words (d = 0.67) and pseudowords (d = 0.60). This was expected because the central focus of phonics programs is upon teaching children to apply the alphabetic system to read novel words. Phonics programs also produced growth in the ability to read irregularly spelled words although the effect size was significantly lower, d = 0.40. This is not surprising because a decoding strategy is less helpful for reading these words. However, alphabetic knowledge is useful for establishing connections in memory that help children read irregular words they have read before. This may explain the contribution of phonics.

Systematic phonics instruction produced significantly greater growth than non-phonics instruction in younger children's reading comprehension ability (d = 0.51). However, the effects of systematic phonics instruction on text comprehension in readers above 1st grade were mixed. Although gains were significant for the subgroup of disabled readers (d = 0.32), they were not significant for the older group in general (d = 0.12).

The conclusion drawn is that growth in word-reading skills is strongly enhanced by systematic phonics instruction when compared to non-phonics instruction for kindergartners and 1st graders as well as for older struggling readers. Growth in reading comprehension is also boosted by systematic phonics instruction for younger students and reading disabled students. Whether growth in reading comprehension is produced generally in students above 1st grade is less clear.

7. Does systematic phonics instruction have an impact on children's growth in spelling?

Systematic phonics instruction produced much growth in spelling among the younger students, that is, kindergartners and 1st graders, d = 0.67, but not among the older students above 1st grade, whose effect size of d = 0.09 did not differ from zero. One factor contributing to the difference is that younger children were given credit for using phonics-based knowledge to produce letter-sound spellings of words as well as correct spellings whereas older children were not. Another factor may be that as children move up in the grades, remembering how to spell words requires knowledge of higher level regularities not covered in systematic phonics programs. A third reason for the poor showing among older students may be that the majority were poor readers who are known to have difficulty learning to spell.

The conclusion drawn is that systematic phonics instruction contributed more than non-phonics instruction in helping kindergartners and 1st graders apply their knowledge of the alphabetic system to spell words. However, it did not improve spelling in students above 1st grade.

8. Is systematic phonics instruction effective with children at different socioeconomic levels?

Systematic phonics instruction helped children at all SES levels make greater gains in reading than did non-phonics instruction. The effect size for low-SES students was d = 0.66, and for middle-class students it was d = 0.44. Both were statistically greater than zero and did not differ from each other. The conclusion drawn is that systematic phonics instruction is beneficial to students regardless of their socioeconomic status.

9. Does the type of control group used to evaluate the effectiveness of systematic phonics instruction make a difference?

The type of nonsystematic or non-phonics instruction given to control groups to evaluate the effectiveness of systematic phonics instruction varied across studies and included the following types: basal programs, regular curriculum, whole language approaches, whole word programs, and miscellaneous programs. The question of whether phonics produced better reading growth than each type of control group was answered affirmatively in each case. The effect sizes were all positive favoring systematic phonics, were all statistically greater than zero, and ranged from d = 0.31 to 0.51. No single effect size differed from any of the others.

The conclusion supported by these findings is that the effectiveness of systematic phonics instruction found in the present meta-analysis did not depend on the type of instruction that students in the control groups received. Students taught systematic phonics outperformed students who were taught a variety of nonsystematic or non-phonics programs, including basal programs, whole language approaches, and whole word programs.

10. Were studies reporting the largest effects of systematic phonics instruction well designed or poorly designed experiments? That is, was random assignment used? Were the sample sizes sufficiently large? Might results be explained by differences between treatment and control groups that existed prior to the experiment rather than by differences produced by the experimental intervention?

The effects of systematic phonics instruction were not diminished when only the best designed experiments were singled out. The mean effect size for studies using random assignment to place students in treatment and control groups, d = 0.45, was essentially the same as that for studies employing quasi-experimental designs, d = 0.43, which utilized existing groups to compare phonics instruction and non-phonics instruction. The mean effect size for studies administering systematic phonics and non-phonics instruction to large samples of students did not differ from studies using the fewest students: for studies using between 80 and 320 students, $\underline{d} = 0.49$; for studies using between 20 and 31 students, $\underline{d} = 0.48$. There were some studies that did not use random assignment and either failed to address the issue of pre-existing differences between treatment and control groups or mentioned that a difference existed but did not adjust for differences in their analysis of results. The effect sizes changed very little when these comparisons were removed from the database, from $\underline{d} = 0.44$ to $\underline{d} = 0.46$.

The conclusion drawn is that the significant effects produced by systematic phonics instruction on children's growth in reading were evident in the most rigorously designed experiments. Significant effects did not arise primarily from the weakest studies.

11. Is enough known about systematic phonics instruction to make recommendations for classroom implementation? If so, what cautions should be kept in mind by teachers implementing phonics instruction?

Findings of the panel regarding the effectiveness of systematic phonics instruction were derived from studies conducted in many classrooms with typical classroom teachers and typical American or English-speaking students from a variety of backgrounds and SES levels. Thus, the results of the analysis are indicative of what can be accomplished when systematic phonics programs are implemented in today's classrooms. Systematic phonics instruction has been used widely over a long period with positive results. A variety of phonics programs have proven effective with children of different ages, abilities, and SES backgrounds. These facts should persuade educators and the public that systematic phonics instruction is a valuable part of a successful classroom reading program. The Panel's findings summarized above serve to illuminate the conditions that make systematic phonics instruction especially effective. However, caution is needed in giving a blanket endorsement to all kinds of phonics instruction.

It is important to recognize that the goals of phonics instruction are to provide children with some key knowledge and skills and to ensure that they know how to apply this knowledge in their reading and writing. Phonics teaching is a means to an end. To be able to make use of letter-sound information, children need phonemic awareness. That is, they need to be able to blend sounds together to decode words, and they need to break spoken words into their constituent sounds to write words. Programs that focus too much on the teaching of letter-sounds relations and not enough on putting them to use are unlikely to be very effective. In implementing systematic phonics instruction, educators must keep the *end* in mind and ensure that children understand the purpose of learning letter-sounds and are able to apply their skills in their daily reading and writing activities.

In addition to this general caution, several particular concerns should be taken into consideration to avoid misapplication of the findings. One concern relates to the commonly heard call for "intensive, systematic" phonics instruction. Usually the term "intensive" is not defined, so it is not clear how much teaching is required to be considered intensive. Questions needing further answers are: How many months or years should a phonics program continue? If phonics has been taught systematically in kindergarten and 1st grade, should it continue to be emphasized in 2nd grade and beyond? How long should single instructional sessions last? How much ground should be covered in a program? That is, how many letter-sound relations should be taught and how many different ways of using these relations to read and write words should be practiced for the benefits of phonics to be maximum? These are among the many questions that remain for future research.

Second, the role of the teacher needs to be better understood. Some of the phonics programs showing large effect sizes are scripted so that teacher judgment is largely eliminated. Although scripts may standardize instruction, they may reduce teachers' interest in the teaching process or their motivation to teach phonics. Thus, one concern is how to maintain consistency of instruction and at the same time encourage unique contributions from teachers. Another concern involves what teachers need to know. Some systematic phonics programs require a sophisticated understanding of spelling, structural linguistics, and word etymology. Teachers who are handed the programs but are not provided with sufficient inservice training to use these programs effectively may become frustrated. In view of the evidence showing the effectiveness of systematic phonics instruction, it is important to ensure that the issue of how best to prepare teachers to carry out this teaching effectively and creatively is given high priority. Knowing that all phonics programs are not the same brings with it the implication that teachers must themselves be educated about how to evaluate different programs and to determine which are based on strong evidence and how they can most effectively use these programs in their own classrooms.

As with any instructional program, there is always the question: "Does one size fit all?" Teachers may be expected to use a particular phonics program with their class, yet it quickly becomes apparent that the program suits some students more than others. In the early grades, children are known to vary greatly in the skills they bring to school. There will be some children who already know most letter-sound correspondences, some children who can even decode words, and others who have little or no letter knowledge. Should teachers proceed through the program and ignore these students? Or should they assess their students' needs and select the types and amounts of phonics suited to those needs? Although the latter is clearly preferable, this requires phonics programs that provide guidance in how to place students into flexible instructional groups and how to pace instruction. However, it is common for many phonics programs to present a fixed sequence of lessons scheduled from the beginning to the end of the school year.

Finally, it is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program. Phonics instruction is never a total reading program. In 1st grade, teachers can provide controlled vocabulary texts that allow students to practice decoding, and they can also read quality literature to students to build a sense of story and to develop vocabulary and comprehension. Phonics should not become the dominant component in a reading program, neither in the amount of time devoted to it nor in the significance attached. It is important to evaluate children's reading competence in many ways, not only by their phonics skills but also by their interest in books and their ability to understand information that is read to them. By emphasizing all of the processes that contribute to growth in reading, teachers will have the best chance of making every child a reader.

Directions for Further Research

Although phonics instruction has been the subject of a great deal of study, there are certain extremely important topics that have received little or no research attention, and there are other topics that, although previously studied, require further research to refine our understanding.

Neglected Topics

Three important but neglected questions are prime candidates for research:

(1) What are the "active ingredients" in effective systematic phonics programs? (2) Is phonics instruction improved when motivational factors are taken into account—not only learners' motivation to learn but also teachers' motivation to teach? (3) How does the use of decodable text as early reading material contribute to the effectiveness of phonics programs?

1. Active Ingredients

Systematic phonics programs—even those of the same type, such as synthetic phonics programs—vary in many respects, as indicated in the Panel's report above. It is important to determine whether some properties are essential and others are not. Because instructional time during the school day is limited, teachers and publishers of beginning reading programs need to know which ingredients of phonics programs yield the most benefit. One example of this line of questions involves the content covered. It is clear that the major letter-sound correspondences, including short and long vowels and digraphs, need to be taught. However, there are other regularities of English as well. How far should instruction extend in teaching all of these potential regularities explicitly? Should children be taught to state regularities, or should emphasis be placed on application in reading and writing activities? To what extent do mnemonic devices such as those used in Jolly Phonics (Lloyd, 1993) and Letterland (Wendon, 1992) speed up the process of learning letter shapes, sounds, and names and facilitate their application in reading and writing? What contribution is made by the inclusion of special markings added to written words to clarify how they should be decoded? Research investigating not only these ingredients of phonics programs but other ingredients as well is needed. These studies should include systematic observation in classrooms to record and analyze the activities of teachers and children using the programs.

2. Motivation

Phonics instruction has often been portrayed as involving "dull drill" and "meaningless worksheets." Such characterizations may accurately describe aspects of some phonics programs, even "effective" ones. Few if any studies have investigated the contribution of motivation to the effectiveness of phonics programs, not only the learner's motivation to learn but also the teacher's motivation to teach. It seems self-evident that the specific techniques and activities used to develop children's letter-sound knowledge and its use in reading and writing should be as relevant and motivating as possible to engage children's interest and attention to promote optimal learning. Moreover, it seems obvious that when the teaching techniques presented to teachers in a phonics program are not only effective but also engaging and enjoyable, teachers will be more successful in their ability to deliver phonics instruction effectively. The lack of attention to motivational factors by researchers in the design of phonics programs is potentially very serious because debates about reading instruction often boil down to concerns about the "relevance" and "interest value" of how something is being taught, rather than the specific content of what is being taught. Future research on phonics instruction should investigate how best to motivate children in classrooms to learn the letter-sound associations and to apply that knowledge to reading and writing. It should also be designed to determine which approaches teachers prefer to use and are most likely to use effectively in their classroom instruction.

3. Decodable Text

Some systematic phonics programs are designed so that children are taught letter-sound correspondences and then provided with little books written carefully to contain the letter-sound relations that were taught. Some programs begin with a very limited set and expand these gradually. The intent of providing books that match children's letter-sound knowledge is to enable them to experience success in decoding words that follow the patterns they know. The stories in such books often involve pigs doing jigs and cats in hats. Other systematic phonics programs make little or no use of decodable books and select the beginning reading material on some other basis. Some educators reject decodable books outright as too stilted and boring. Surprisingly, very little research has attempted to determine whether the use of decodable books in systematic phonics programs has any influence on the progress that some or all children make in learning to read.

Other Important Topics

The findings of the Panel indicated that systematic phonics instruction provides beginning readers, at-risk readers, disabled readers, and low-achieving readers with a substantial edge in learning to read over alternative forms of instruction not focusing at all or only incidentally on the alphabetic system. However, studies in the database were insufficient in number or in design to address several important satellite questions about the effects of phonics instruction.

Some programs teach many letter-sound relations before children begin using them while other programs introduce a few and then provide reading and writing activities that allow children to apply the correspondences they have learned right away. The latter approach would appear to be preferable, but is it? In what ways does earlier application facilitate growth in reading and writing?

Programs differ in how much time is consumed teaching alphabetic knowledge and word-reading skills. It is unclear how long phonics instruction should continue through the grades. A few studies in the Panel's database indicated that large effect sizes were produced and maintained in the 2nd and 3rd years of instruction for children who were at risk for future reading problems and who began receiving systematic phonics instruction in kindergarten or 1st grade (Blachman et al., (1999); Brown & Felton, 1990; Torgesen et al., 1999). See Table 4 (Appendix E). This suggests that systematic phonics instruction should extend from kindergarten to 2nd grade, but the question remains whether additional instruction will produce further benefits.

It will also be critical to objectively determine the ways in which systematic phonics instruction can be optimally incorporated and integrated in complete and balanced programs of reading instruction. Part of this effort should be directed at preservice and inservice education to provide teachers with decisionmaking frameworks to guide their selection, integration, and implementation of phonics instruction within a complete reading program.

Another line of questions for research centers around older children above 1st grade who have acquired some reading ability but are reading substantially below grade level. When systematic phonics instruction is introduced to these children, do they have difficulty acquiring alphabetic knowledge and decoding strategies because they have already learned other ways to process print that undermine the acquisition and incorporation of these new processes into their reading? If so, perhaps special steps are required to address this problem. A related question is how can systematic phonics instruction be made more effective for low-achieving readers who have below-average intelligence as well as reading problems. Perhaps instruction in decoding needs to be combined with instruction in reading comprehension strategies to remediate their reading problems.

When systematic phonics instruction is introduced to children who have already acquired some reading skill as a result of another program that does not emphasize phonics, one wonders about the impact of attempting to teach students new strategies when old tricks have already been learned. Findings of the Panel indicated that the impact of systematic phonics instruction was much reduced among children who were introduced to it presumably for the first time in 2nd grade and above. (This presumption may not be accurate, however, because most studies did not state what kind of instruction children had already experienced.) Additional research is needed to study how systematic phonics instruction is received by children who are already reading; whether there are sources of conflict; and, if so how to address them instructionally. A related question is whether the sequence of instruction makes a difference. It may be that children do better when a year of systematic phonics instruction precedes a year of whole language instruction than when the reverse is the case.

References

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Anderson, R. C., Hiebert, E. F., Wilkinson, I. A. G., & Scott, J. (1985). Becoming a nation of readers. Champaign, IL: Center for the Study of Reading.

Aukerman, R. C. (1971). Approaches to beginning reading. New York: John Wiley & Sons.

Aukerman, R. C. (1981). The basal reader approach to reading. New York: John Wiley & Sons.

Aukerman, R. C. (1984). Approaches to beginning reading (2nd ed.). New York: John Wiley & Sons.

Balmuth, M. (1982). The roots of phonics: A historical introduction. New York: McGraw-Hill.

Barr, R., Kamil, M., Mosenthal, P., & Pearson, P. (1991). Handbook of reading research (Vol. 2). New York: Longman.

Baumann, J., Hoffman, J., Moon, J., & Duffy-Hester, A. (1998). Where are teachers' voices in the phonics/whole language debate? Results from a survey of U.S. elementary classroom teachers. The Reading Teacher, 51, 636-650.

Bear, D. R., Invernizz, M., Templeton, S., & Johnson, F. (2000). Words their way: Word study for phonics, vocabulary and spelling instruction. Upper Saddle River, NJ: Merrill.

Beck, I., & Mitroff, D. (1972). The rationale and design of a primary grades reading system for an individualized classroom. Pittsburgh: University of Pittsburgh Learning Research and Development Center.

Blachman, B. (in press). Phonological awareness. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), Handbook of reading research (Vol. 3). Mahwah, NJ: Erlbaum.

Bond, G., & Dykstra, R. (1967/1998). The cooperative research program in first grade reading. Reading Research Quarterly, 2, 5-142.

Brown, I., & Felton, R. (1990). Effects of instruction on beginning reading skills in children at risk for reading disability. Reading and Writing: An Interdisciplinary Journal, 2, 223-241.

Bruck, M. (1993). Component spelling skills of college students with childhood diagnoses of dyslexia. Learning Disability Quarterly, 16, 171-184.

Campbell, D., & Stanley, J. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Carnine, D., & Silbert, J. (1979). Direct instruction reading. Columbus, OH: Merrill.

Chall, J. S. (1967). Learning to read: The great debate (1st ed.). New York, NY: McGraw-Hill.

Chall, J. S. (1983). Learning to read: The great debate (2nd ed.). New York, NY: McGraw-Hill.

Chall, J. (1992). The new reading debates: Evidence from science, art, and ideology. Teachers College Record, 94, 315-328.

Chall, J. (1996a). Learning to read: The great debate (revised, with a new foreword). New York: McGraw-Hill.

Chall, J. (1996b). Stages of reading development (2nd ed.). Fort Worth, TX: Harcourt-Brace.

Chall, J., & Feldmann., S. (1966). First grade reading: An analysis of the interactions of professed methods, teacher implementation, and child background. The Reading Teacher, 19, 569-575.

Clay, M. M. (1993). Reading recovery: A guidebook for teachers in training. Portsmouth, NH: Heinemann.

Cohen, J. (1988). Statistical power analysis for the behavior sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cox, A. R. (1991). Structures and techniques: Multisensory teaching of basic language skills. Cambridge, MA: Educators Publishing Service.

Dahl, K. L., Sharer, P. L., Lawson, L. L., & Grogran, P. R. (1999). Phonics instruction and student achievement in whole language first-grade classrooms. Reading Research Quarterly, 34, 312-341.

Dickson, S. (1972a). Sing, spell, read, & write. Chesapeake, VA: International Learning Systems.

Dykstra, R. (1968). The effectiveness of code- and meaning-emphasis in beginning reading programs. The Reading Teacher, 22, 17-23.

Ehri, L. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), Handbook of reading research (Vol. 2) (pp. 383-417). New York: Longman.

Ehri, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), Reading acquisition (pp. 107-143). Hillsdale, NJ: Erlbaum.

Ehri, L. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed. pp. 323-358). Newark, DE: International Reading Association.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), Word recognition in beginning literacy (pp. 3-40). Mahwah, NJ: Erlbaum.

Ehri, L., Deffner, N., & Wilce, L. (1984). Pictorial mnemonics for phonics. Journal of Educational Psychology, 76, 880-893.

Engelmann, S. (1980). Direct instruction. Englewood Cliffs, NJ: Prentice-Hall.

Engelmann, S., & Bruner, E. (1969). Distar reading program. Chicago: SRA.

Engelmann, S., & Bruner, E. (1978). DISTAR reading: An instructional system. Chicago: Science Research Associates.

Engelmann, S., & Bruner, E. (1988). Reading mastery I/II fast cycle: Teacher's guide. Chicago: Science Research Associates, Inc.

Engelmann, S., & Osborn, J. (1987). DISTAR Language I. Chicago: Science Research Associates.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), Word recognition in beginning literacy (pp. 3-40). Mahwah, NJ: Erlbaum.

Ehri, L., Deffner, N., & Wilce, L. (1984). Pictorial mnemonics for phonics. Journal of Educational Psychology, 76, 880-893.

Engelmann, S. (1980). Direct instruction. Englewood Cliffs, NJ: Prentice-Hall.

Engelmann, S., & Bruner, E. (1969). Distar reading program. Chicago: SRA.

Engelmann, S., & Bruner, E. (1978). DISTAR reading: An instructional system. Chicago: Science Research Associates.

Engelmann, S., & Bruner, E. (1988). Reading mastery I/II fast cycle: Teacher's guide. Chicago: Science Research Associates, Inc.

Engelmann, S., & Osborn, J. (1987). DISTAR Language I. Chicago: Science Research Associates.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), Word recognition in beginning literacy (pp. 3-40). Mahwah, NJ: Erlbaum.

Ehri, L., Deffner, N., & Wilce, L. (1984). Pictorial mnemonics for phonics. Journal of Educational Psychology, 76, 880-893.

Engelmann, S. (1980). Direct instruction. Englewood Cliffs, NJ: Prentice-Hall.

Engelmann, S., & Bruner, E. (1969). Distar reading program. Chicago: SRA.

Engelmann, S., & Bruner, E. (1978). DISTAR reading: An instructional system. Chicago: Science Research Associates.

Engelmann, S., & Bruner, E. (1988). Reading mastery I/II fast cycle: Teacher's guide. Chicago: Science Research Associates, Inc.

Engelmann, S., & Osborn, J. (1987). DISTAR Language I. Chicago: Science Research Associates.

Foorman, B., Francis, D., Fletcher, J., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. Journal of Educational Psychology, 90, 37-55.

Freppon, P. A., & Dahl, K. L. (1991). Learning about phonics in a whole language classroom. Language Arts, 68, 190-197.

Freppon, P. A., & Headings, L. (1996). Keeping it whole in whole language: A first grade teacher's instruction in an urban whole language classroom. In E. McIntyre & M. Pressley (Eds.), Balanced instruction: Strategies and skills in whole language (pp. 65-82). Norwood, MA: Christopher-Gordon.

Gaskins, I., Downer, M., & Gaskins, R. (1986). Introduction to the Benchmark school word identification/vocabulary development program. Media, PA: Benchmark School.

Gilllingham, A., & Stillman, B. (1940). Remedial training for children with specific disability in reading, spelling, and penmanship. New York, NY: Sackett & Williams Lithographing Co.

Gillingham, A., & Stillman, B. W. (1979). Remedial training for children with specific disability in reading, spelling, and penmanship. Cambridge, MA: Educators Publishing Service.

Goodman, K. S. (1993). Phonics phacts. Portsmouth, NH: Heinemann.

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. Remedial and Special Education, 7, 6-10.

Greaney, K., Tunmer, W., & Chapman, J. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. Journal of Educational Psychology, 89, 645-651.

Grundin, H. U. (1994). If it ain't whole, it ain't language—or back to the basics of freedom and dignity. In F. Lehr & J. Osborn (Eds.), Reading, language, and literacy (pp. 77-88). Mahwah, NJ: Erlbaum.

Harris, T., & Hodges, R. (Eds.). (1995). The literacy dictionary. Newark, DE: International Reading Association.

Haskell, D., Foorman, B., & Swank, P. (1992). Effects of three orthographic/phonological units on first-grade reading. Remedial and Special Education, 13, 40-49.

Holdaway, D. (1979). The foundations of literacy. Sydney: Ashton-Scholastic.

Hoover, W., & Gough, P. (1990). The simple view of reading. Reading and Writing, 2, 127-160.

Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery Program. Journal of Educational Psychology, 85, 112-126.

Johnson, B. (1989). DSTAT: Software for the meta-analytic review of research literatures. Hillsdale, NJ: Erlbaum.

Juel, C. (1994). Learning to read and write in one elementary school. New York: Springer-Verlag.

Kameenui, E. J., Simmons, D. C., Chard, D., & Dickson, S. (1997). Direct instruction reading. In S. A. Stahl & D. A. Hayes (Eds.), Instructional models in reading (pp. 59-84). Mahwah, NJ: Erlbaum.

Klesius, J., Griffith, P., & Zielonka, P. (1991). A whole language and traditional instruction comparison: Overall effectiveness and development of the alphabetic principle. Reading Research and Instruction, 30, 47-61.

Leach, D., & Siddall, S. (1990). Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. British Journal of Educational Psychology, 60, 349-355.

Lindamood, C. H., & Lindamood, P. C. (1984). Auditory discrimination in depth. Austin, TX: PRO-ED, Inc.

Lloyd, S. (1993). The phonics handbook. UK: Jolly Learning.

Levin, H., & Williams, J. (Eds.). (1970). Basic studies on reading. New York: Basic Books.

Lovett, M., Borden, S., DeLuca, T., Lacerenza, L., Benson, N., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. Developmental Psychology, 30, 805-822.

Lovett, M., Lacerenza, L., Borden, S., Frijters, J., Steinbach, K., & DePalma, M. (in press). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. Journal of Educational Psychology.

Lovett, M., & Steinbach, K. (1997). The effectiveness of remedial programs for reading disabled children of different ages: Does the benefit decrease for older children? Learning Disability Quarterly, 20, 189-210.

Lovett, M., Warren-Chaplin, P., Ransby, M., & Borden, S. (1990). Training the word recognition skills of reading disabled children: Treatment and transfer effects. Journal of Educational Psychology, 82, 769-780.

Lovitt, T. C., & DeMier, D. M. (1984). An evaluation of the Slingerland method with LD youngsters. Journal of Learning Disabilities, 17, 267-272.

McCracken, G., & Walcutt, C. (1963, 1965, 1966, 1975). Basic reading. Philadelphia, PA: Lippincott.

McKenna, M., Stahl, S., & Reinking, D. (1994). A critical commentary on research, politics, and whole language. Journal of Reading Behavior, 26, 211-233.
Lovitt, T. C., & DeMier, D. M. (1984). An evaluation of the Slingerland method with LD youngsters. Journal of Learning Disabilities, 17, 267-272.

McCracken, G., & Walcutt, C. (1963, 1965, 1966, 1975). Basic reading. Philadelphia, PA: Lippincott.

McKenna, M., Stahl, S., & Reinking, D. (1994). A critical commentary on research, politics, and whole language. Journal of Reading Behavior, 26, 211-233.

Meyer, L. A. (1983). Increased student achievement in reading: One district's strategies. Research in Rural Education, 1, 47-51.

Mills, H., O'Keefe, T., & Stephens, D. (1992). Looking closely: Exploring the role of phonics in one whole language classroom. Urbana, IL: National Council of Teachers of English.

Morris, D. (1992). Case studies in teaching beginning readers: The Howard Street tutoring manual. Boone, NC: Fieldstream.

Morris, D., & Perney, J. (1984). Developmental spelling as a predictor of first grade reading achievement. Elementary School Journal, 84, 441-457.

Ogden, S., Hindman, S., & Turner, S. (1989). Multisensory programs in the public schools: A brighter future for LD children. Annals of Dyslexia, 39, 247-267.

Pearson, P., Barr, R., Kamil, M., & Mosenthal, P. (1984). Handbook of reading research. New York: Longman.

Popp, H. M. (1975). Current practices in the teaching of beginning reading. In J. B. Carroll & J. S. Chall (Eds.), Toward a literate society (pp. 101-146). New York: McGraw-Hill.

Purves, A. (Ed.). (1994). Encyclopedia of English studies and language arts. New York: Scholastic.

Rack, J., Snowling, M., & Olson, R. (1992). The nonword reading deficit in developmental dyslexia: A review. Reading Research Quarterly, 27, 29-53.

Routman, R. (1996). Literacy at the crossroads. Portsmouth, NH: Heinemann.

Santa, C., & Hoien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. Reading Research Quarterly, 34, 54-79.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanaugh & I. G. Mattingly (Eds.), Language by eye and by ear (pp. 293-317). Cambridge, MA: MIT Press.

Share, D. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. Cognition, 55, 151-218.

Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading achievement. Journal of Educational Psychology, 76, 1309-1324.

Snider, V. (1990). Direct instruction reading with average first-graders. Reading Improvement, 27, 143-148.

Stahl, S. A. (1999). Why innovations come and go: The case of whole language. Educational Researcher.

Stahl, S. A., Duffy-Hester, A. M., & Stahl, K. A. D. (1998). Everything you wanted to know about phonics (but were afraid to ask). Reading Research Quarterly, 35, 338-355.

Stahl, S. A., & Miller, P. D. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. Review of Educational Research, 59 (1), 87-116.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 21, 360-406.

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves rading and spelling in inner-city second language learners. British Journal of Educational Psychology, 69, 587-605.

Tangel, D., & Blachman, B. (1995). Effect of phoneme awareness instruction on the invented spelling of first-grade children: A one-year follow-up. Journal of Reading Behavior, 27, 153-185.

Taylor, D. (1998). Beginning to read & the spin doctors of science: The political campaign to change America's mind about how children learn to read. Urbana, IL: National Council of Teachers of English.

Torgesen, J., Wagner, R., Rashotte, C., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology, 91, 579-593.

Tunmer, W., & Chapman, J. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme & R. Joshi (Eds.), Reading and spelling: Development and disorders (pp. 33-67). Mahwah, NJ: Erlbaum.

Tunmer, W., & Hoover, W. (1993). Phonological recoding skill and beginning reading. Reading and Writing: An Interdisciplinary Journal, 5, 161-179.

Vandervelden, M., & Siegel, L. (1997). Teaching phonological processing skills in early literacy: A developmental approach. Learning Disability Quarterly, 20, 63-81.

Vickery, K., Reynolds, V., & Cochran, S. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. Annals of Dyslexia, 37, 189-200.

Weaver, C. (1998). Experimental research: On phonemic awareness and on whole language. In C. Weaver (Ed.), Reconsidering a balanced approach to reading (pp. 321-371). Urbana, IL: National Council of Teachers of English.

Wendon, L. (1992). First steps in Letterland. Cambridge, UK: Letterland Ltd.

Woodcock, R. (1987). Woodcock reading mastery tests - revised. Circle Pines, MN: American Guidance Service.

PART II: PHONICS INSTRUCTION

Appendices

Appendix A

Studies Included in the Meta-Analysis

Note: Studies were assigned numbers during the screening process. Numbers missing in the list were assigned to studies that were rejected for the analysis. These are listed separately.

62 Barr, R. (1972). The influence of instructional conditions on word recognition errors. Reading Research Quarterly, 7 (3), 509-529.

63 Barr, R. (1974). The effect of instruction on pupil reading strategies. Reading Research Quarterly, 10, (4), 555-582.

03 Blachman, B., Tangel, D., Ball, E., Black, R. & McGraw, D. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. Reading and Writing: An Interdisciplinary Journal, 11, 293-273.

61 Blachowicz, C., McCarthy, L., & Ogle, D. (1979). Testing phonics: A look at children's response biases. Illinois School Research and Development, 16 (1), 1-6.

04 Bond, C., Ross, S., Smith, L., & Nunnery, J. (1995-96). The effects of the sing, spell, read, and write program on reading achievement of beginning readers. Reading Research and Instruction, 35, 122-141.

05 Brown, I., & Felton, R. (1990). Effects of instruction on beginning reading skills in children at risk for reading disability. Reading and Writing: An Interdisciplinary Journal, 2, 223-241.

64 Carnine, D. (1977). Phonics versus look-say: Transfer to new words. The Reading Teacher, 30, 636-640.

71 Carnine, D. (1980). Phonic versus whole-word correction procedures following phonic instruction. Education and Treatment of Children, 3(4), 323-329.

08 Eldredge, L. (1991). An experiment with a modified whole language approach in first-grade classrooms. Reading Research and Instruction, 30, 21-38.

09 Evans, M., & Carr, T. (1985). Cognitive abilities, conditions of learning, and the early development of reading skill. Reading Research Quarterly, 20, 327-350.

11 Foorman, B., Francis, D., Fletcher, J., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. Journal of Educational Psychology, 90, 37-55.

12 Foorman, B., Francis, D., Novy, D., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. Journal of Educational Psychology, 83, 456-469.

13 Foorman, B., Francis, D., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. (1997). Early interventions for children with reading disabilities. Scientific Studies of Reading, 1, 255-276.

59 Freppon, P. (1991). Children's concepts of the nature and purpose of reading in different instructional settings. Journal of Reading Behavior, 23, 139-163.

15 Fulwiler, G., & Groff, P. (1980). The effectiveness of intensive phonics. Reading Horizons, 21, 50-54.

72 Gersten, R., Darch, C., & Gleason, M. (1988). Effectiveness of a direct instruction academic kindergarten for low-income students. The Elementary School Journal, 89, 227-240.

66 Gersten, R., Keating, T., & Becker, W. (1988). The continued impact of the direct instruction model:

Longitudinal studies of follow through students. Education and Treatment of Children, 11 (4), 318-327.

68 Gillon, G., & Dodd, B. (1997). Enhancing the phonological processing skills of children with specific reading disability. European Journal of Disorders of Communication, 32, 67-90.

17 Gittelman, R., & Feingold, I. (1983). Children with reading disorders—I. Efficacy of reading remediation. Journal of Child Psychology and Psychiatry and Allied Disciplines, 24, 167-191.

18 Greaney, K., Tunmer, W., & Chapman, J. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. Journal of Educational Psychology, 89, 645-651.

60 Griffith, P., Klesius, J., & Kromey, J. (1992). The effect of phonemic awareness on the literacy development of first grade children in a traditional or a whole language classroom. Journal of Research in Childhood Education, 6, 85-92.

22 Haskell, D., Foorman, B., & Swank, P. (1992). Effects of three orthographic/phonological units on first-grade reading. Remedial and Special Education, 13, 40-49.

73 Kameenui, E., Stein, M., Carnine, D., & Maggs, A. (1981). Primary level word attack skills based on isolated word, discrimination list and rule application training. Reading Education, 6 (2), 46-55.

26 Klesius, J., Griffith, P., & Zielonka, P. (1991). A whole language and traditional instruction comparison: Overall effectiveness and development of the alphabetic principle. Reading Research and Instruction, 30, 47-61.

28 Leach, D., & Siddall, S. (1990). Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. British Journal of Educational Psychology, 60, 349-355.

29 Leinhardt, G., & Engel, M. (1981). An iterative evaluation of NRS: Ripples in a pond. Evaluation Review, 5, 579-601.

75 Lovett, M., Lacerenza, L., Borden, S., Frijters, J., Steinbach, K., & DePalma, M. (in press). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. Journal of Educational Psychology.

32 Lovett, R., Ransby, M., Hardwick, N., Johns, M., & Donaldson, S. (1989). Can dyslexia be treated? Treatment-specific and generalized treatment effects in dyslexic children's reponse to remediation. Brain and Language, 37, 90-121.

33 Lovett, M., & Steinbach, K. (1997). The effectiveness of remedial programs for reading disabled children of different ages: Does the benefit decrease for older children? Learning Disability Quarterly, 20, 189-210.

34 Lovett, M., Warren-Chaplin, P., Ransby, M., & Borden, S. (1990). Training the word recognition skills of reading disabled children: Treatment and transfer effects. Journal of Educational Psychology, 82, 769-780.

35 Lum, T., & Morton, L. (1984). Direct instruction in spelling increases gain in spelling and reading skills. Special Education in Canada, 58, 41-45.

36 Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the SEARCH/TEACH tutoring approach with middle-class students at risk for reading failure. Elementary School Journal, 92, 573-586.

37 Marston, D., Deno, S., Kim, D., Diment, K., & Rogers, D. (1995). Comparison of reading intervention approaches for students with mild disabilities. Exceptional Children, 62, 20-37.

38 Martinussen, R., & Kirby, J. (1998). Instruction in successive and phonological processing to improve the reading acquisition of at-risk kindergarten children. Developmental Disabilities Bulletin, 26, 19-39.

41 Oakland, T., Black, J., Stanford, G., Nussbaum, N., & Balise, R. (1998). An evaluation of the dyslexia training program: A multisensory method for promoting reading in students with reading disabilities. Journal of Learning Disabilities, 31, 140-147.

65 Peterson, M., & Haines, L. (1992). Orthographic analogy training with kindergarten children: Effects on analogy use, phonemic segmentation, and letter-sound knowledge. Journal of Reading Behavior, 24 (1), 109-127.

44 Santa, C., & Hoien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. Reading Research Quarterly, 34, 54-79.

47 Silberberg, N., Iversen, I., & Goins, J. (1973). Which remedial reading method works best? Journal of Learning Disabilities, 6, 18-27.

48 Snider, V. (1990). Direct instruction reading with average first-graders. Reading Improvement, 27, 143-148.

74 Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves rading and spelling in inner-city second language learners. British Journal of Educational Psychology, 69, 587-605.

67 Torgesen, J., Wagner, R., & Rashotte, C. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. Scientific Studies of Reading, 1 (3), 217-234.

51 Torgesen, J., Wagner, R., Rashotte, C., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology, 91, 579-593.

52 Traweek, K., & Berninger, V. (1997). Comparisons of beginning literacy programs: Alternative paths to the same learning outcome. Learning Disability Quarterly, 20, 160-168.

53 Tunmer, W., & Hoover, W. (1993). Phonological recoding skill and beginning reading. Reading and Writing: An Interdisciplinary Journal, 5, 161-179.

69 Umbach, B., Darch, C., & Halpin, G. (1989). Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. Journal of Instructional Psychology, 16, 23-30.

54 Vandervelden, M., & Siegel, L. (1997). Teaching phonological processing skills in early literacy: A developmental approach. Learning Disability Quarterly, 20, 63-81.

55 Vickery, K., Reynolds, V., & Cochran, S. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. Annals of Dyslexia, 37, 189-200.

70 Wise, B., Olson, R., Anstett, M., Andrews, L., Terjak, M., Schneider, V., Kostuch, J., & Driho, L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real-world issues. Behavior Research Methods, Instruments and Computers, 21 (2), 173-180.

57 Wilson, K., & Norman, C. (1998). Differences in word recognition based on approach to reading instruction. Alberta Journal of Educational Research, 44, 221-230.

Appendix B

List of Studies in the Original Database That Were Excluded

01 Becker, W., & Gersten, R. (1982). A follow-up of follow through: The later effects of the direct instruction model on children in fifth and sixth grades. American Educational Research Journal, 19(1), 75-92.

02 Berninger, V., Vaughan, K., Abbott, R., Brooks, A., Abbott, S., Rogan, L., Reed, E., & Graham, S. (1998). Early intervention for spelling problems: Teaching functional spelling units of varying size with a multiple-connections framework. Journal of Educational Psychology, 90, 587-605.

06 Eddowes, E. (1990). Teaching reading in kindergarten: Two contrasting approaches. Reading Improvement, 27, 220-223.

07 Ehri, L., & Wilce; L. (1987). Cipher versus cue reading: An experiment in decoding acquisition. Journal of Educational Psychology, 79 (1), 3-13.

10 Fielding-Barnsley, R (1997). Explicit instruction in decoding benefits children high in phonemic awareness and alphabet knowledge. Scientific Studies of Reading, 1, 85-98.

14 Foorman, B., & Liverman, D. (1989). Visual and phonological processing of words: A comparison of good and poor readers. Journal of Learning Disabilities, 22, 349-355.

16 Gettinger, M. (1987). Prereading skills and achievement under three approaches to teaching word recognition. Journal of Research and Development in Education, 19, 1-9.

19 Guyer, B., Banks, S., & Guyer, K. (1993). Spelling improvement for college students who are dyslexic. Annals of Dyslexia, 43, 186-193.

20 Guyer, B., & Sabatino, D. (1989). The effectiveness of a multisensory alphabetic phonetic approach with college students who are learning disabled. Journal of Learning Disabilities, 22, 430-434.

21 Hall, D., & Cunningham, P. (1988). Context as a polysyllabic decoding strategy. Reading Improvement, 25, 261-264.

23 Jeffrey, W., & Samuels, S. (1967). Effect of method of reading training on initial learning and transfer. Journal of Verbal Learning and Verbal Behavior, 6, 354-358.

24 Kaiser, S., Palumbo, K., Bialozor, R., & McLaughlin, T. (1989). The effects of direct instruction with rural remedial students: A brief report. Reading Improvement, 26, 88-93.

25 Kibby, M. (1979). The effects of certain instructional conditions and response modes on initial word learning. Reading Research Quarterly, 15, 147-171.

27 Kuder, S. (1990). Effectiveness of the DISTAR Reading Program for children with learning disabilities. Journal of Learning Disabilities, 23, 69-71.

30 Lovett, M., Borden, S., DeLuca, T., Lacerenza, L., Benson, N., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. Developmental Psychology, 30, 805-822.

31 Lovett, M., Ransby, M., & Barron, R. (1988). Treatment, subtype, and word type effects in dyslexic children's response to remediation. Brain and Language, 34, 328-349.

39 McNeil, J., & Donant, L. (1980). Transfer effect of word recognition strategies. Journal of Reading Behavior, 12(2), 97-103.

40 Meyer, L. (1982). The relative effects of word-analysis and word-supply correction procedures with poor readers during word-attack training. Reading Research Quarterly, 17(4), 544-555.

42 Olson, R., & Wise, B. (1992). Reading on the computer with orthographic and speech feedback. Reading and Writing: An Interdisciplinary Journal, 4, 107-144.

43 Olson, R., Wise, B., Ring, J., & Johnson, M. (1997). Computer-based remedial training in phoneme awareness and phonological decoding: Effects on the post-training development of word recognition. Scientific Studies of Reading, 1(3), 235-253.

45 Scherzer, C. & Goldstein, D. (1982). Children's first reading lesson: Variables influencing within-lesson emotional behavior and post-lesson achievement. Journal of Educational Psychology, 74(3), 382-392.

46 Schwartz, S. (1988). A comparison of componential and traditional approaches to training reading skills. Applied Cognitive Psychology, 2, 189-201.

49 Spiegel, D., Fitzgerald, J., & Heck, M. (1985). Teaching children to use a context-plus-phonics strategy. Reading Horizons, 25, 176-185.

50 Sullivan, H., Okada, M., & Niedermeyer, F. (1971). Learning and transfer under two methods of word-attack instruction. American Educational Research Journal, 8(2), 227-239.

56 Walton, P. (1995). Rhyming ability, phoneme identify, letter-sound knowledge, and the use of orthographic analogy by pre-readers. Journal of Educational Psychology, 87(4), 587-597.

58 Wise, B., & Olson, R. (1992). How poor readers and spellers use interactive speech in a computerized spelling program. Reading and Writing: An Interdisciplinary Journal, 4, 145-163.

Appendix C

In the reports of experiments included in the meta-analysis examining the effects of phonics instruction, references were supplied for the programs and material used to teach systematic phonics. These are listed below.

Synthetic Phonics Programs

03 Blachman et al., 1999

Blachman, B. (1987). An alternative classroom reading program for learning disabled and other low achieving children. In R. Bowler (Ed.), Intimacy with language: A forgotten basic in teacher education (pp. 49-55). Baltimore, MD: Orton Dyslexia Society.

Engelmann, S. (1969). Preventing failure in the primary grades. Chicago: Science Research Associates.

Slingerland, B. (1971). A multi-sensory approach to language arts for specific language disability children: A guide for primary teachers. Cambridge, MA: Educators Publishing Service.

Primary phonics series – published by Educator's Publishing Service.

Selected stories from Scott Foresman basal reading series (none of its other materials was used).

04 Bond et al., 1995

Dickson, S. (1972a). Sing, spell, read, & write. Chesapeake, VA: International Learning Systems.

05 Brown & Felton, 1990

Lippincott basic reading. (1981). Riverside, NJ: Macmillan Publishing Company.

11 Foorman et al., 1998

Open Court Reading. (1995). Collections for young scholars. Chicago and Peru, IL: SRA/McGraw-Hill.

12 Foorman et al., 1991

Scott, Foresman (1985). Scott, Foresman reading. Glenview, IL: Author.

Modern Curriculum Press. (1986). Phonics practice readers, series B. Cleveland, OH: Author.

Elwell, C., Murray, P., & Kucia, M. (1988). Modern Curriculum Press phonics program. Cleveland, OH: Modern Curriculum Press.

13 Foorman et al., 1997

Cox, A. R. (1991). Structures and techniques: Multisensory teaching of basic language skills. Cambridge, MA: Educators Publishing Service.

Gillingham, A., & Stillman, B. (1979). Remedial training for children with specific disability in reading, spelling, and penmanship. Cambridge, MA: Educators Publishing Service.

Fulwiler & Groff, 1980

McCracken, G., & Walcutt, C. (1975). Basic reading. Philadelphia: Lippincott.

17 Gittelman & Feingold, 1983

Pollack, C. (1967a). Phonic readiness kit. Brooklyn, NY: Book-Lab, Inc.

Pollack, C. (1967b). The intersensory reading program: My own reading book and my own writing book, part I. Brooklyn, NY: Book-Lab, Inc.

Pollack, C., & Atkins, R. (1971). The intersensory reading program: My own reading book, part II. Brooklyn, NY: Book-Lab, Inc.

Pollack, C., & Lane, P. (1969). The lipreader, part I. Brooklyn, NY: Book-Lab, Inc.

Pollack, C., & Lane, P. (1971) The lipreader, part II. Brooklyn, NY: Book-Lab, Inc.

28 Leach & Sidall, 1990

Engelmann, S. (1980). Direct instruction. Englewood Cliffs, NJ: Prentice-Hall.

Engelmann, S., & Bruner, E. (1975). DISTAR reading I & II: An instructional system. Chicago: Science Research Associates.

Engelmann, S., Haddox, P., & Bruner, E. (1983). Teach your child to read in 100 easy lessons. New York: Simon & Schuster.

29 Leinhardt & Engel, 1981

Beck, I. (1977). Comprehension during the acquisition of decoding skills. In J. T. Guthrie (Ed.), Cognition, curriculum, and comprehension. Newark, DE: International Reading Association.

Beck, I., & Mitroff, D. (1972). The rationale and design of a primary grades reading system for an individualized classroom. Pittsburgh: University of Pittsburgh Learning Research and Development Center.

Lovett & Steinbach, 1997

Engelmann, S., & Bruner, E. C. (1988). Reading mastery I/II fast cycle: Teacher's guide. Chicago: Science Research Associates, Inc.

Engelmann, S., Carnine, L., & Johnson, G. (1988). Corrective reading, word attack basics, decoding A. Chicago: Science Research Associates, Inc.

Engelmann, S., Johnson, G., Carnine, L., Meyer, L., Becker, W., & Eisele, J. (1988). Corrective reading: Decoding strategies, decoding B1. Chicago: Science Research Associates, Inc.

37 Marston et al., 1995

Engelmann, S. (1980). Direct instruction. Englewood Cliffs, NJ: Educational Technology Publications.

Science Research Associates (1988). Reading mastery. New York: SRA, Macmillan/McGraw-Hill.

38 Martinussen & Kirby, 1998

Das, J., Mishra, R., & Pool, J.(1995). An experiment on cognitive remediation of reading difficulties. Journal of Learning Disabilities, 28, 66-79.

41 Oakland et al., 1998

Beckham, P. B., & Biddle, M. L. (1989). Dyslexia training program books. Cambridge, MA: Educators Publishing Service.

48 Snider, 1990

Carnine, D., & Silbert, J. (1979). Direct instruction reading. Columbus, OH: Merrill.

51 Torgesen et al., 1999

Lindamood, C. H., & Lindamood, P. C. (1984). Auditory discrimination in depth. Austin, TX: PRO-ED, Inc.

Short stories from:

Hannah, S. (1993). Early literacy series, I & II. Edmonton, Alberta, Canada: The Centre for Literacy.

Smith, L. (1992) Auditory discrimination reading series-sequence B. Independence, MO: Poppin Creations.

52 Traweek & Berninger, 1997

Engelmann, S., & Bruner, E. C. (1978). DISTAR reading: An instructional system. Chicago: Science Research Associates.

55 Vickery et al., 1987

Gillingham, A., & Stillman, B. W. (1956). Remedial training for children with specific disability in reading, spelling, and penmanship (5th ed). Cambridge, MA: Educators Publishing Service.

Cox, A. (1984). Structures and techniques: Multisensory teaching of basic language skills. Cambridge, MA: Educators Publishing Service.

Waites, L., & Cox, A. (1969). Developmental language disability . . . Basic training . . . Remedial language training. Cambridge, MA: Educators Publishing Service.

69 Umbach et al., 1992

Reading mastery series (1986)

Carnine, D., & Silbert, J. (1979). Direct instruction reading. Columbus, OH: Merrill.

72 Gersten et al., 1988

Becker, W., Engelmann, S., Carnine, D., & Rhine, R. (1981). Direct instruction models. In R. Rhine (Ed.), Encouraging change in American schools: A decade of experimentation (pp. 95-154). New York: Academic Press.

Carnine, D., Carnine, L., Karp, J., & Weisbert, P. (1988) Kindergarten for economically disadvantaged students: The direct instruction component. In C. Wargen (Ed.), A resource guide to public school early childhood programs (pp. 73-98). Alexandria, VA: ASCD

Engelmann, S., & Osborn, J. (1987). DISTAR language I. Chicago: Science Research Associates.

74 Stuart, 1999

Lloyd, S. (1992). The phonics handbook. UK: Jolly Learning.

Lovett et al., in press

Engelmann, S., & Bruner, E. C. (1988). Reading mastery I/II fast cycle: Teacher's guide. Chicago: Science Research Associates, Inc.

Engelmann, S., Carnine, L., & Johnson, G. (1988). Corrective reading, word attack basics, decoding A. Chicago: Science Research Associates, Inc.

Engelmann, S., Johnson, G., Carnine, L., Meyer, L., Becker, W., & Eisele, J. (1988). Corrective reading: Decoding strategies, decoding B1. Chicago: Science Research Associates, Inc.

Phonics Programs Emphasizing Larger Phonological Subunits

11 Foorman et al., 1998

Hiebert, E., Colt, J., Catto, S., & Gary, E. (1992). Reading and writing of first-grade students in a restructured chapter 1 program. American Educational Research Journal, 29, 545-572.

13 Foorman et al., 1997

Traub, N., & Bloom, F. (1992). Recipe for reading (3rd ed.). Cambridge, MA: Educators Publishing Service.

33 Lovett & Steinbach, 1997

Gaskins, I., Downer, M., & Gaskins, R. (1986). Introduction to the Benchmark school word identification/vocabulary development program. Media, PA: Benchmark School.

Gaskins, I., Downer, M., Anderson, R., et al. (1988). A metacognitive approach to phonics: Using what you know to decode what you don't know. Remedial and Special Education, 9, 36-41, 66.

44 Santa & Hoien, 1999

Morris, D., Shaw, B., & Perney, J. (1990). Helping low readers in grades 2 and 3: An after school volunteer program. Elementary School Journal, 91, 133-150.

Santa, C. (1998). Early steps: Learning from a reader. Kalispell, MT: Scott.

51 Torgesen et al., 1999

Short stories from HBJ bookmark Series

53 Tunmer & Hoover, 1993

Clay , M. (1985). The early detection of reading difficulties. Auckland: Heinemann.

Clay, M. (1991). Becoming literate: The construction of inner control. Auckland: Heinemann.

Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery program. Journal of Educational Psychology, 85, 112-126.

75 Lovett et al., in press

Gaskins, I., Downer, M., & Gaskins, R. (1986). Introduction to the Benchmark school word identification/vocabulary development program. Media, PA: Benchmark School.

Miscellaneous Phonics Programs:

34 Lovett et al., 1990

Dixon, R., Engelmann, S., & Meier, M. (1980). Spelling mastery, a direction instruction series - Level A. Chicago: Science Research Asssociates.

Dixon, R., & Engelmann, S. (1981). Spelling mastery, a direction instruction series - Level B. Chicago: Science Research Associates.

Mantzicopoulos et al., 1992

Cradler, J., Bechthold, D., & Bechthold, C. (1974). Success-controlled optimal reading experience (SCORE). Hillsborough, CA: Learning Guidance Systems.

Makar, B. (1985). Primary phonics. Cambridge, MA: Educators Publishing Service.

Phonics practice readers. (1986). Cleveland, OH: Modern Curriculum Press.

Hall & Price (1984). Explode the code. Cambridge, MA: Educators Publishing Service.

Bechthold, C., & Bechthold, D. (1976). Signs for sounds. Hillsborough, CA. Educational Support Systems.

APPENDIX D

Table 2

Treatment-Control Group Comparisons in the Database Grouped by Type of Phonics Program and Coded for Instructional Unit, Grade, Reading Ability of Participants, Length of Treatment, Type of Control Group, and Overall Effect size on Literacy Outcome Measures

Identity/Type[a] of Program	Inst. Unit	Grade/ Abil.[b]	Length[c]	Control[d]	d[e]
EMPHASIS ON SYNTHETIC PHONICS (S)					
74 Jolly Phonics (S)	Class	K at risk	12 wks.	Big Books (WL)	0.73
38 Successive phonics (S)	Sm gp	K at risk	8 wks.	Reg. curr.	0.62
03 Blachman PA (S)	Sm gp	K at risk	2-3 yrs.	Basal	0.72/0.36
04 SingSpellReadWrite (S)	Class	K	1 yr.	Basal	0.51
51 Lindamood PA (S)	Tutor	K at risk	3 yrs	Reg. curr.	0.33/0.67
72 Direct Instruction (S)	Class	K at risk	4 yrs.	Reg. curr.	-- /0.24
29 NRS-2 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.45
29 NRS-3 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.44
29 NRS-4 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.33
29 NRS-6 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.70
12 Synthetic basal (S)	Class	1st	1 yr	Whole word	2.27
04 SingSpellReadWrite (S)	Class	1st	1 yr.	Basal	0.25
15 Lippincott (S)	Class	1st	1 yr.	Whole word	0.84
48 Direct Instruction (S)	Sm gp	1st	1 yr.	Basal/Prev. yr.	0.38[f]
28 Direct Instruction (S)	Tutor	1st	10 wks.	Misc. (child reads)	1.99
08 Modif. Whole Lang (S)	Class	1st at risk	1 yr.	Basal	0.63
11 Open Court (S)	Class	1st at risk	1 yr.	Whole language	0.91
52 Direct Instruction (S)	Class	1st at risk	1 yr.	Whole language	0.07
69 Direct Instruction (S)	Sm gp	1st at risk	1 yr.	Basal	1.19
05 Lippincott (S)	Sm gp	1st at risk	2 yrs.	Whole word	0.48/.52
72 Direct Instruction (S)	Class	1st at risk	3 yrs.	Reg. curr.	--/0.00
04 SingSpellReadWrite (S)	Class	2nd	1 yr.	Basal	0.38
57 Sequential phonics (S)	Class	2nd	1 yr.	Whole language	-0.47
11 Open Court (S)	Class	2nd lo ach.	1 yr.	Whole language	0.12
37 Direct Instruction (S)	Class	gr 1-6 lo ach	10 wks.	Reg. curr.	0.01
55 Orton-Gillingham (S)	Class	3rd	1 yr.	Previous prog. (RC)	0.04
55 Orton-Gillingham (S)	Class	4th	1 yr.	Previous prog. (RC)	0.04
55 Orton-Gillingham (S)	Class	5th	1 yr.	Previous prog. (RC)	0.61
55 Orton-Gillingham (S)	Class	6th	1 yr.	Previous prog. (RC)	0.43
33 Lovett Dir. Inst. (S)	Sm gp	gr 2-3 RD	9 wks(35 s)	Misc. (Study skills)	0.24
33 Lovett Dir. Inst. (S)	Sm gp	gr 4 RD	9 wks(35 s)	Misc. (Study skills)	1.42
33 Lovett Dir. Inst. (S)	Sm gp	gr 5-6 RD	9 wks(35 s)	Misc. (Study skills)	0.09
17 Intersensory method (S)	Tutor	age 7-13 RD	18 wks.	Misc. (Subj. tutor)	0.53

Table 2 (Continued)

75 Lovett Dir. Inst. (S)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.24
32 Decoding skills (S)	Sm gp	age 8-13 RD	40 sessions	Misc. (Study skills)	0.39
47 Lippincott (S)	Sm gp	3rd RD	1 yr.	Whole word	0.50
13 Orton-Gillingham (S)	Sm gp	gr 2-3 RD	1 yr.	Whole word	0.27
41 Orton-Gillingham (S)	Sm gp	M=11yr RD	2 yrs.	Reg. curr.	--0.54
47 Orton-Gillingham (S)	Sm gp	3rd RD	1 yr.	Whole word	0.04
55 Orton-Gillingham (S)	Class	3rd lo ach.	1 yr.	Previous prog. (RC)	0.63
55 Orton-Gillingham (S)	Class	4th lo ach.	1 yr.	Previous prog. (RC)	0.19
55 Orton-Gillingham (S)	Class	5th lo ach.	1 yr.	Previous prog. (RC)	-0.20
55 Orton-Gillingham (S)	Class	6th lo ach.	1 yr.	Previous prog. (RC)	0.13

Table 2 (Continued)

Identity/Type[a] of Program	Inst. Unit	Grade/ Abil.[b]	Length[c]	Control[d]	d[e]
EMPHASIS ON BLENDING LARGER SUBUNITS AS WELL AS PHONEMES (LU)					
51 Embedded (LU)	Tutor	K at risk	3 yrs.	Reg. curr.	0.32/0.17
11 Embedded (LU)	Class	1st at risk	1 yr.	Whole language	0.36
11 Embedded (LU)	Class	2nd lo ach	1 yr.	Whole language	0.03
13 Onset-rime (LU)	Sm gp	gr 2-3 RD	1 yr.	Whole word	-0.11
44 RRD[g]-Early Steps (LU)	Tutor	1st at risk	1 yr.	Whole language	0.76
53 RRD[g]-Phonograms (LU)	Tutor	1st at risk	42 sessions	Reg. curr.	3.71
18 RRD[g]-Rime anal.(LU)	Tutor	gr 2-5 lo ach	11 wks.	Whole word	0.37
33 Lovett Analogy (LU)	Sm gp	gr 2/3 RD	9 wks (35 s)	Misc. (Study skills)	0.49
33 Lovett Analogy (LU)	Sm gp	gr 4 RD	9 wks (35 s)	Misc. (Study skills)	1.41
33 Lovett Analogy (LU)	Sm gp	gr 5/6 RD	9 wks (35 s)	Misc. (Study skills)	-0.25
75 Lovett Analogy (LU)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.50
COMBINATION PROGRAMS (C)					
75 Dir.Inst.+Analogy.(C)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.60
75 Analogy+Dir.Inst. (C)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.21
MISCELLANEOUS PHONICS (M)					
54 Developmental (M)	Sm gp	K at risk	12 wk.	Reg. curr. extended	0.47
09 Traditional basal (M)	Class	1st	1 yr.	Whole language	0.60
22 Analyze phonemes (M)	Sm gp	1st	6 wks.	Whole word	-0.07
22 Analyze onset-rimes (M)	Sm gp	1st	6 wks.	Whole word	0.14
26 Traditional basal (M)	Class	1st	1 yr.	Whole language	0.20
59 Sequential phonics (M)	Class	1st	1 yr/less	Whole language	0.00
60 Traditional basal (M)	Class	1st	1 yr.	Whole language	-0.33
36 Phonetic read/spell (M)	Tutor	1st at risk	1 yr (50 s)	Reg. curr.	0.53
35 Spelling mastery (M)	Class	2nd	1 yr.	Tradit. spell (RC)	0.38
34 Analytic (M)	Sm gp	age 7-13 RD	9 wks (35 s)	Misc. (Study skills)	0.16

[a] The programs listed as Direct Instruction include Reading Mastery and DISTAR.

[b] Information about grade/reading ability refers to the point in time when instruction began. RD refers to children classified as reading disabled. Lo ach refers to children above first grade who were identified as low achievers in their ability to read. At risk refers to kindergartners or first graders who performed poorly either on reading tests or on tests predictive of poor reading. If not marked, the sample consisted of normally developing readers.

[c] s refers to the number of sessions.

[d] RC means regular curriculum. WL means whole language. Misc. means miscellaneous category.

[e] Effect sizes listed singly are those observed at the end of training that lasted one year or less. When training lasted longer than one year, the first effect size reports the outcome at the end of the first year and the second effect size reports the outcome at the end of training.

[f] This effect size was not measured immediately after training but following a delay of six months.

[g] RRD refers to a program derived from Reading Recovery that was modified to include systematic phonics instruction in which phonemes were taught along with larger phonological units such as onsets, rimes and spelling patterns.

APPENDIX E

Table 3

Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Whether Effect Sizes Were Homogeneous at $p < 0.05$, and Whether Effect Sizes Differed From Each Other at $p < 0.05$. Effect Sizes Refer to Outcomes Immediately After Training or At the End of One School Year, Whichever Came First, Unless Labeled as Followup or End of Training.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Time of Posttest					
End of Training	65	0.41*	No	0.36 to 0.47	n.s.
End of Training or One Year[a]	62	0.44*	No	0.38 to 0.50	
Followup	7	0.28*	Yes	0.10 to 0.46	
End of Training[b]	6	0.51*	Yes	0.32 to 0.70	n.s.
Followup	6	0.27*	Yes	0.07 to 0.46	
Outcome Measures					
Decoding regular words	30	0.67*	No	0.57 to 0.77	DecR = DecP;
Decoding pseudowords	40	0.60*	No	0.52 to 0.67	Both >
Reading misc. words	59	0.40*	No	0.34 to 0.46	RW, Spel,
Spelling words	37	0.35*	No	0.28 to 0.43	Oral,
Reading text orally	16	0.25*	No	0.15 to 0.36	Comp.
Comprehending text	35	0.27*	No	0.19 to 0.36	
Characteristics of Participants					
Grade					
Kind. & First	30	0.55*	No	0.47 to 0.62	K-1st >
2nd-6th, RD	32	0.27*	Yes	0.18 to 0.36	2nd-6th/RD
Younger Grades					
Kindergarten	7	0.56*	Yes	0.40 to 0.73	
First Grade	23	0.54*	No	0.46 to 0.63	
Kindergarten and First Graders on Outcome Measures					
Decoding regular words	8	0.98*	No	0.81 to 1.16	DecR >
Decoding pseudowords	14	0.67*	No	0.56 to 0.78	RMW, Co, Or;
Reading misc. words	23	0.45*	No	0.37 to 0.53	Spel > Or;
Spelling words	13	0.67*	No	0.54 to 0.79	DecP > Or.
Reading text orally	6	0.23*	No	0.05 to 0.41	
Comprehending text	11	0.51*	No	0.36 to 0.65	
2nd-6th, RD on Outcome Measures					
Decoding regular words	17	0.49*	No	0.34 to 0.65	DecR > Sp;
Decoding pseudowords	13	0.52*	Yes	0.37 to 0.66	DecP >
Reading misc. words	23	0.33*	No	0.22 to 0.44	Sp,Co.
Spelling words	13	0.09ns	Yes	-0.04 to 0.23	
Reading text orally	6	0.24*	Yes	0.08 to 0.39	
Comprehending text	11	0.12ns	Yes	-0.04 to 0.28	

Table 3 (Continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Grade and Reading Ability					
Kindergarten At Risk	6	0.58*	Yes	0.40 to 0.77	1AR > 2N, 2AR, RD
1st Normal	14	0.48*	No	0.38 to 0.58	
1st At Risk	9	0.74*	No	0.56 to 0.91	
2nd-6th Normal	7	0.27*	Yes	0.12 to 0.43	
2nd-6th Lo Achievers	8	0.15ns	Yes	-0.06 to 0.36	
Reading Disabled	17	0.32*	Yes	0.18 to 0.46	
Socioeconomic Status					
Low SES	6	0.66*	Yes	0.48 to 0.85	n.s.
Middle SES	10	0.44*	No	0.28 to 0.60	
Varied	14	0.37*	Yes	0.26 to 0.48	
Not Given	32	0.43*	No	0.34 to 0.51	
Characteristics of Instruction					
Type of Phonics Program					
Synthetic	39	0.45*	No	0.39 to 0.52	n.s
Larger Phon. Units	11	0.34*[d]	No	0.16 to 0.52	
Miscellaneous	10	0.27*	Yes	0.08 to 0.46	
Specific Phonics Programs					
NRS-Beck LRDC (S)	4	0.47*	Yes	0.33 to 0.60	n.s.
Direct Instruction (S)	4	0.48*	No	0.13 to 0.83	
Lovett Direct Instruct (S)	4	0.41*	Yes	0.04 to 0.77	
Lovett Analogy (LU)	4	0.48*	Yes	0.11 to 0.86	
Lippincott (S)	3	0.68*	Yes	0.43 to 0.93	
Orton Gillingham (S)	10	0.23*	Yes	0.06 to 0.39	
Sing Spell Read Write (S)	3	0.35*	Yes	0.21 to 0.50	
Synthetic Phonics For Various Readers Groups					
K & 1st At Risk[c]	9	0.64*	Yes	0.49 to 0.80	K&1AR > 2-6LA, 2-6N
1st Normal	8	0.54*	No	0.43 to 0.65	
2nd-6th Normal	6	0.27*	Yes	0.11 to 0.43	
2nd-6th Lo Achievers	6	0.14ns	Yes	-0.10 to 0.39	
Reading Disabled	9	0.36*	Yes	0.18 to 0.54	
Unit of Instruction					
Tutor	8	0.57*[d]	No	0.38 to 0.77	n.s.
Small Group	27	0.43*	Yes	0.34 to 0.52	
Class	27	0.39*	No	0.31 to 0.48	
Type of Control Group					
Basal	10	0.46*	Yes	0.37 to 0.55	n.s.
Regular Curriculum	16	0.41*	No	0.27 to 0.54	
Whole Language	12	0.31*	No	0.16 to 0.47	
Whole Word	10	0.51*	No	0.35 to 0.67	
Miscellaneous	14	0.46*	Yes	0.28 to 0.63	

Table 3 (Continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen	95% CI	Contrasts
Characteristics of the Design of Studies					
Assignment of Participants to Treatment and Control Groups					
Random	23	0.45*	Yes	0.32 to 0.58	n.s.
Nonequivalent Groups	39	0.43*	No	0.37 to 0.50	
Sample Size					
20 to 31	14	0.48*	No	0.26 to 0.70	n.s.
32 to 52	16	0.31*	Yes	0.15 to 0.47	
53 to 79	16	0.36*	No	0.23 to 0.49	
80 to 320	16	0.49*	No	0.41 to 0.57	

---

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

[a] Effect sizes indicate literacy outcomes at the end of training for studies lasting 1 year or less, and at the end of the first school year for studies that continued training beyond 1 year.

[b] The six studies in both comparisons were the same studies.

[c] The kindergarten and 1st grade at-risk groups had identical ds and were combined.

[d] This effect size was adjusted to reduce the impact of one atypically large outlier.

Table 4

Characteristics of Sets of Studies of Special Interest

Type of Program[a]	Inst. Unit	Grade/ Abil.	Length	Control	d
STUDIES WITH TRAINING LASTING MORE THAN A YEAR[c]					
03 Blachman PA (S)	Sm gp	K at risk	2-3 yrs	Basal	0.72/0.64/0.36
51 Lindamood PA (S)	Tutor	K at risk	3 yrs	Reg. curr.	0.33/0.75/0.67
51 Embedded (LU)	Tutor	K at risk	3 yrs	Reg. curr.	0.32/0.28/0.17
05 Lippincott (S)	Sm gp	1st at risk	2 yrs	Whole word	0.48/0.52
72 Direct Instruction (S)	Class	K at risk	4 yrs	Reg. curr.	--/0.24
72 Direct Instruction (S)	Class	1st at risk	3 yrs	Reg. curr.	--/0.00
41 Orton-Gillingham (S)	Sm gp	M=11yr RD	2 yrs	Reg. curr.	--/0.54
STUDIES MEASURING IMMEDIATE OUTCOMES AND LONG-TERM OUTCOMES[b]					
18 Rime analogy (LU)	Tutor	gr 2-5 lo ach	11 wks	Whole word	0.37/0.56 (1 yr.)
36 Phonetic read/spell (M)	Tutor	1st at risk	50 ses	Reg. curr.	0.53/0.32 (1 yr.)
44 Early Steps (LU)	Tutor	1st at risk	1 yr	Whole language	0.76/0.86 (4 mo.)
47 Orton-Gillingham (S)	Sm gp	3rd RD	1 yr	Whole word	0.04/-0.47 (6 mo.)
47 Lippincott (S)	Sm gp	3rd RD	1 yr	Whole word	0.50/0.33 (6 mo.)
48 Direct Instruction (S)	Sm gp	1st	1 yr	Basal (Prev. yr)	--/0.38 (6 mo.)
74 Jolly Phonics (S)	Class	K at risk	12 wks	Big Books (WL)	0.73/0.28 (1 yr.)
2ND-6TH LOW ACHIEVERS					
11 Embedded (LU)	Class	2nd lo ach	1 yr	Whole language	0.03
11 Open Court (S)	Class	2nd lo ach.	1 yr	Whole language	0.12
18 Rime analogy (LU)	Tutor	gr 2-5 lo ach	11 wks	Whole word	0.37
37 Direct Instruction (S)	Class	gr 1-6 lo ach	10 wks	Reg. curr.	0.01
55 Orton-Gillingham (S)	Class	3rd lo ach.	1 yr	Previous prog. (RC)	0.63
55 Orton-Gillingham (S)	Class	5th lo ach.	1 yr	Previous prog. (RC)	-0.20
55 Orton-Gillingham (S)	Class	6th lo ach.	1 yr	Previous prog. (RC)	0.13
55 Orton-Gillingham (S)	Class	4th lo ach.	1 yr	Previous prog. (RC)	0.19
TUTORING COMPARISONS					
51 Lindamood PA (S)	Tutor	K at risk	3 yrs	Reg. curr. (class)	0.33/0.67
51 Embedded (LU)	Tutor	K at risk	3 yrs	Reg. curr. (class)	0.32/0.17
28 Direct Instruction (S)	Tutor	1st	10 wks	Misc. (child reads) (tutor)	1.99
36 Phonetic read/spell (M)	Tutor	1st at risk	50 ses	Reg. curr. (class)	0.53
44 Early Steps (LU)	Tutor	1st at risk	1 yr	Whole lang. (sm gp)	0.76
53 Phonograms (LU)	Tutor	1st at risk	42 ses	Reg. curr. (class)	3.71
17 Intersensory method (S)	Tutor	age 7-13 RD	18 wks	Misc. (Subj. tutor)	0.53
18 Rime analogy (LU)	Tutor	gr 2-5 lo ach	11 wks	Whole word (tutor)	0.37

---

[a] Letters in parentheses refer to the type of phonics program: S (synthetic), LU (Larger subunits), M (Miscellaneous).

[b] The first effect size is for the immediate posttest and the second is for the delayed posttest. The length of the delay between posttests is given in parentheses.

[c] When 3 effect sizes are reported, these refer to effects at the end of each year of training.

Table 5

Number of Comparisons by Grade and Reading Ability

Grade	Reading Ability			
	Normally Developing	At Risk/ Low Achievers	Reading Disabled	Total
Kindergarten	1	6 (K-AR)	--	7
First Grade	14 (1N)	9 (1-AR)	--	23
Second Grade	3 (2-6N)	2 (2-6 AR)	--	5
3rd-6th Grades	4 (2-6N)	4 (2-6 AR)	6 (RD)	14
Mixed grades	--	2 (2-6 AR)	11 (RD)	13
Total	22	23	17	62

Note. The symbols in parentheses refer to the groups that were created for the meta-analysis.

Appendix F

Table 6

Characteristics of the Treatment-Control Group Comparisons Utilizing Specific Phonics Programs That Were Included in the Meta-AnalysiS

Identify/Type of Program	Inst. Unit	Grade/ Abil.	Length	Control	d
28 Direct Instruction (S)	Tutor	1st	10 wks.	Misc. (child reads)	1.99
52 Direct Instruction (S)	Class	1st at risk	1 yr.	Whole language	0.07
69 Direct Instruction (S)	Sm gp	1st at risk	1 yr.	Basal	1.19
37 Direct Instruction (S)	Class	gr 1-6 lo ach	10 wks.	Reg. curr.	0.01
33 Lovett Dir. Inst. (S)	Sm gp	gr 4 RD	9 wks(35 s)	Misc. (Study skills)	1.42
33 Lovett Dir. Inst. (S)	Sm gp	gr 2-3 RD	9 wks(35 s)	Misc. (Study skills)	0.24
33 Lovett Dir. Inst. (S)	Sm gp	gr 5-6 RD	9 wks(35 s)	Misc. (Study skills)	0.09
75 Lovett Dir. Inst. (S)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.24
33 Lovett Analogy (LU)	Sm gp	gr 4 RD	9 wks (35 s)	Misc. (Study skills)	1.41
33 Lovett Analogy (LU)	Sm gp	gr 2/3 RD	9 wks (35 s)	Misc. (Study skills)	0.49
33 Lovett Analogy (LU)	Sm gp	gr 5/6 RD	9 wks (35 s)	Misc. (Study skills)	-0.25
75 Lovett Analogy (LU)	Sm gp	age 6-13 RD	70 hrs	Misc.(Study+Math)	0.50
15 Lippincott (S)	Class	1st	1 yr.	Whole word	0.84
05 Lippincott (S)	Sm gp	1st at risk	2 yrs.	Whole word	0.48
47 Lippincott (S)	Sm gp	3rd RD	1 yr.	Whole word	0.50
29 NRS-6 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.70
29 NRS-4 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.33
29 NRS-3 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.44
29 NRS-2 (Beck) (S)	Sm gp	1st	1 yr.	Basal	0.45
55 Orton-Gillingham (S)	Class	3rd	1 yr.	Previous prog. (RC)	0.04
55 Orton-Gillingham (S)	Class	4th	1 yr.	Previous prog. (RC)	0.04
55 Orton-Gillingham (S)	Class	5th	1 yr.	Previous prog. (RC)	0.61
55 Orton-Gillingham (S)	Class	6th	1 yr.	Previous prog. (RC)	0.43
55 Orton-Gillingham (S)	Class	3rd lo ach.	1 yr.	Previous prog. (RC)	0.63
55 Orton-Gillingham (S)	Class	4th lo ach.	1 yr.	Previous prog. (RC)	0.19
55 Orton-Gillingham (S)	Class	5th lo ach.	1 yr.	Previous prog. (RC)	-0.20
55 Orton-Gillingham (S)	Class	6th lo ach.	1 yr.	Previous prog. (RC)	0.13
13 Orton-Gillingham (S)	Sm gp	gr 2-3 RD	1 yr.	Whole word	0.27
47 Orton-Gillingham (S)	Sm gp	3rd RD	1 yr.	Whole word	0.04
04 SingSpellReadWrite (S)	Class	K	1 yr.	Basal	0.51
04 SingSpellReadWrite (S)	Class	1st	1 yr.	Basal	0.25
04 SingSpellReadWrite (S)	Class	2nd	1 yr.	Basal	0.38

Appendix F (continued)

Table 7

Descriptions of the Specific Phonics Programs Examined in the Meta-Analysis

1. Direct Instruction. The Direct Instruction program is based on a behavioral analysis of the steps involved in learning to decode (Carnine & Silbert, 1979; Engelmann, 1980; Engelmann & Bruner, 1969, 1978, 1988; Engelmann & Osborn, 1987; Kameenui et al., 1997). At the beginning of the program, students are not taught letter names but only letter-sound relations through highly structured instruction that uses cueing and reinforcement procedures derived from a behavioral analyses of instruction. The task of decoding is broken down into its component parts, and each of these parts is taught separately, from letter sounds to blending to reading words in context. Instruction is scripted and the lessons are fast paced, with high student participation. The text for the first-year program is written in a script that, although it preserves English spelling, contains printed marks that cue the reader about silent letters and different vowel sounds. Children practice in specially constructed books containing taught sounds, although children may be encouraged to read widely in children's literature as well (e.g., Meyer, 1983).

2. Lovett Direct Instruction. The synthetic phonics program used by Lovett and Steinbach (1997) and Lovett et al. (in press) adopts the Direction Instruction model to remediate the decoding and phonemic awareness difficulties of severely disabled readers. Children are taught phonological analysis and blending (phonemic awareness) orally and also letter sound associations in the context of word recognition and decoding instruction. The program focuses on training sound blending and acquisition of a left-to-right phonological decoding strategy. The special orthography highlights salient features of many letters and provides visual cues such as symbols over long vowels, letter size variations, and connected letters to facilitate learning. Cumulative, systematic review and many opportunities for overlearning are hallmarks of this approach. New material is not introduced until the child fully masters previously instructed material.

3. Lovett Analogy. A second program also used with severely disabled readers by Lovett and Steinbach (1997) and Lovett et al. (in press) was adapted from the Benchmark Word Identification/Vocabulary Development program developed by Gaskins et al. (1986). This program is strongly metacognitive in its focus. It teaches children how to use four metacognitive strategies to decode words: reading words by analogy, detecting parts of words that are known, varying the pronunciations of vowels to maintain flexibility in decoding attempts, and "peeling off" prefixes and suffixes in words. Children learn a set of 120 key words exemplifying high-frequency spelling patterns, 5 words per day. They learn to segment the words into subunits so that they can use these known words and their parts to read other similarly spelled words. They learn letter-sound associations for vowels and affixes. Various types of texts provide children with practice applying the strategies taught.

4. Lippincott. The Lippincott Basic Reading Series (McCracken & Walcutt, 1963, 1975) is a direct code method which, from the outset, approaches reading from a phonic/linguistic perspective. Beginning with children's spoken language, the Lippincott program teaches in a systematic manner how to use the alphabetic code to move from printed words to oral language. Instruction begins with short-a and builds knowledge of regular sound/symbol relationships. Children are first taught to decode phonetically regular words, with blending of phonic elements directly taught. Once they are proficient, long vowels and irregular spellings are introduced. Although the primary instructional focus is on decoding, another goal of this method is the instant recognition of words. However, rather than relying on a "context clue" approach to word recognition, children are taught how and why the letters come to

Appendix F (continued)

Table 7 (continued)

represent these words, and they learn to "break the code" to decipher new words independently. Review and reinforcement are an integral part of the program. Spelling is sometimes taught as one component of the reading lesson with spelling lists developed from the words introduced in each unit of reading instruction (Brown & Felton, 1990).

5. NRS by Beck and Mitroff. The New Primary Grades Reading System for an Individualized Classroom (NRS) was developed by Beck and Mitroff (1972). It is a code-breaking approach. The program begins by teaching self-management skills, letter-sound correspondences, and chain blending to decode words. Children are taught to pronounce the first letter of a word followed by the second letter and then to blend the two sounds; then they pronounce the third letter and add it to the blend. In the first lesson, children are taught five isolated letter-sound relations, and once they are known, children are immediately taught to blend them to form real words. Subsequent letter-sounds are taught one at a time and blended with the earlier letters. Not only synthetic phonics but also analytic phonics is taught as children explore words and their parts. The method is linguistic as well because the major spelling patterns of words are displayed in texts to draw attention to similarities and contrasts, and because there is minimum teaching of explicit pronunciation rules. Instruction is individualized. After the first two levels, children work through the curriculum at different rates.

6. Orton Gillingham. The Orton-Gillingham approach (Cox, 1991; Gillingham & Stillman, 1979) begins with the direct teaching of individual letters paired with their sounds using a Visual-Auditory-Kinesthetic-Tactile (VAKT) procedure that involves tracing the letter while saying its name and sound, blending letters together to read words and sentences, and finally reading short stories constructed to contain only taught sounds. Spelling words from dictation is also part of an Orton-Gillingham lesson. Each letter-sound is learned to mastery through repetition. More advanced lessons involve teaching learners to blend syllables together and read more complex texts. Among those approaches based on Orton and Gillingham's work are the Slingerland approach (Lovitt & DeMier, 1984), the Spaulding Approach, Recipe for Reading, and Alphabetic Phonics (Ogden, Hindman, & Turner, 1989). There are differences among these approaches, largely in the sequencing of materials, but they all have the general characteristics discussed.

7. Sing, Spell, Read & Write. The Sing, Spell, Read and Write (SSRW) program (Dickson, 1972) also teaches synthetic phonics. It consists of several charts, books (both readers and workbooks), letter and word cards, tests, and audio tapes. The tapes contain songs about several phonics generalizations. Through the tapes, the students learn the sounds of letters and letter combinations. Also songs combined with charts help students learn the spellings of words. The lessons begin by teaching letter-sounds in isolation for each letter of the alphabet. When students have mastered certain sounds, they begin reading phonetic storybooks. The first five books each focus on a different vowel sound. The remaining books expand the vocabulary in a way that is consistent with the letter-sounds taught. Students are taught to spell the words they learn to read, with the words presented in sentences. Most of the writing students do involves filling in blanks or answering questions related to words being learned. The program has a "racetrack" which is posted in classrooms and notes students' progress by placement of a race car on the chart (Bond et al., 1995-96).

Appendix G

Studies in the Phonics Database, Their Characteristics, and Effect Sizes

(*Note:* key to this chart is on page 2-176)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
03 - Blachman et al., (1999)	.	.	.			.	.	.	.	.	.	.	.	.	.	.	.	.	.
Blachman PA	Syn	Basal	SmG	2-3yrs(41s-,20m/d)	K	AR	Low	NE	No	159	Imm.	0.72	-0.17	1.08	0.94	.	1.04	.	.
Blachman PA				(1st gr=30 m/d)						128	2nd yr tr.	0.64	0.35	0.81	0.53	.	0.86	.	.
Blachman PA				(2nd gr=30 m/d)						106	3rd yr tr.	0.36	0.42	0.55	0	.	0.45	.	.
04 - Bond et al., 1995																			
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr.(20 lessons)	K	N	Var	NE	No	144	Imm.	0.51	0.38	.	.	.	1.01	0.13	.
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr.	1st	N	Var	NE	No	276	Imm.	0.25	0.23	.	0.14	.	0.6	0.03	.
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr.	2nd	N	Var	NE	No	320	Imm.	0.38	0.44	.	0.18	.	0.55	0.33	.
05 - Brown & Felton, 1990																			
Lippincott	Syn	Wh.W.	SmG	2 yrs.	1st	AR	NG	R	No	47	Imm.	0.48	0.02	.	0.51	.	0.92	.	.
Lippincott											2nd yr tr.	0.52	0.51	0.63	0.38	.	0.55	.	.
08 - Eldredge, 1991																			
Modified Whole Language	Syn	Basal	Class	1 yr. (15m/d)	1st	AR	Low	NE	No	105	Imm.	0.63	.	.	.	0.83	0.43	.	.
09 - Evans, 1985																			
Traditional Basal	Misc	Wh.L.	Class	1 yr.	1st	N	Var	NE	NG	20*(N=-247)	Imm.	0.6	.	.	.	0.6	.		

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
11 - Foorman et al., 1998																			
Open Court	Syn	Wh.L.	Class	1 yr. (30m/d)	1st	AR	Var	NE	NG	68	Imm.	0.91	1.63	1.14	0.56	0.32			
Embedded	LU	Wh.L.	Class	1 yr.	1st	AR	Var	NE	NG	70	Imm.	0.36	0.56	0.51	0.26	0.1			
Open Court	Syn	Wh.L.	Class	1 yr.	2nd	LA	Var	NE	NG	35	Imm.	0.12	0.52	0.32	-0.19	-0.19			
Embedded	LU	Wh.L.	Class	1 yr.	2nd	LA	Var	NE	NG	57	Imm.	0.03	0.37	0.22	-0.25	-0.24			
12 - Foorman et al., 1991																			
Synthetic basal	Syn	Wh.W.	Class	1 yr. (45 m/d)	1st	N	Mid	NE	No	6*(N=8-0)	Imm.	2.27	1.92	2.67	2.21				
13 - Foorman et al., 1997																			
Orton-Gillingham	Syn	Wh.W.	SmG	1 yr. (60 m/d)	gr 2-3	RD	Mid	NG	Yes	67	Imm.	0.27	0.17	0.58	0.05				
Onset-rime	LU	Wh.W.	SmG	1 yr.	gr 2-3	RD	Mid	NG	Yes	85	Imm.	-0.11	-0.19	0.09	-0.23				
15 - Fulwiler & Groff, 1980																			
Lippincott	Syn	Wh.W.	Class	1 yr.	1st	N	NG	NE	NG	147	Imm.	0.84		0.91		0.76			
17 - Gittelman & Feingold, 1983																			
Intersensory Method	Syn	Misc.	Tutor	18 wks.(54s)	7-13yr	RD	Mid	R	No	56	Imm.	0.53	0.76	0.67	0.12	0.57			
18 - Greaney et al., 1997																			
RRD-Rime analogy	LU	Wh.W.	Tutor	11 wks(31s.3-0m)	gr 2-5	LA	NG	R	No	36	Imm.	0.37	0.39				0.51	0.2	
RRD-Rime analogy										34	follow up	0.56	0.47				0.76	0.44	

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year. Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
22 - Haskell et al., 1992																			
Analyze Onset-Rimes	Misc	Wh.W.	SmG	6 wks(15s. 20m)	1st	N	Mid	R	No	24	Imm.	0.14	0.2	0.09					
Analyze Phonemes	Misc	Wh.W.	SmG	6 wks(15s. 20m)	1st	N	Mid	R	No	24	Imm.	-0.07	-0.08	-0.06					
26 - Klesius et al., 1991																			
Traditional Basal	Misc	Wh.L	Class	1 yr	1st	N	Var	NE	Yes	6*(N=1-12)	Imm.	0.2			0.36	0.18	0.07		
28 - Leach & Siddall, 1990																			
Direct Instruction	Syn	Misc	Tutor	10 wks (15m/d)	1st	N	NG	R	No	20	Imm.	1.99				1.8		2.18	
29 - Leinhardt & Engel, 1981																			
NRS-study 2 (Beck)	Syn	Basal	SmG	1 yr.	1st	N	NG	NE	Yes	187	Imm.	0.45	0.45						
NRS-study 3 (Beck)	Syn	Basal	SmG	1 yr.	1st	N	NG	NE	Yes	263	Imm.	0.44	0.44						
NRS-study 4 (Beck)	Syn	Basal	SmG	1 yr.	1st	N	NG	NE	Yes	256	Imm.	0.33	0.33						
NRS-study 6 (Beck)	Syn	Basal	SmG	1 yr.	1st	N	NG	NE	Yes	241	Imm.	0.7	0.7						
32 - Lovett et al., 1989																			
Decoding Skills	Syn	Misc.	SmG	40 ses (33-40h)	8-13yr	RD	Mid	R	No	118	Imm.	0.39	0.78	0.7	0.42	0.07	0.1	0.27	
33 - Lovett & Steinbach, 1997																			
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 2/3	RD	NG	R	No	28	Imm.	0.49	-0.12	0.85			0.75		
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 4	RD	NG	R	No	22	Imm.	1.41	0.84	2.06			1.33		

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	**Type of Phonics**	**Control Group**	**Tr.unit**	**Length of Training**	**Grade/ Age**	**Reading Ability**	**SES**	**Group Assign.**	**Sig Pre-test Diff**	**Total N**	**Time of Post-test**	**Mean**	**Word ID**	**Dec**	**Spell**	**Comp**	**Nonw**	**Oral Read**	**Gen. Read**
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 5/6	RD	NG	R	No	24	Imm.	-0.25	-0.49	-0.15			-0.1		
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 2/3	RD	NG	R	No	32	Imm.	0.24	0.02	0.24			0.46		
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 4	RD	NG	R	No	25	Imm.	1.42	1.03	1.53			1.7		
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 5/6	RD	NG	R	No	27	Imm.	0.09	-0.24	0.25			0.25		
34 - Lovett et al., 1990																			
Analytic	Misc	Misc.	SmG	9wks (35h)	7-13yr	RD	Mid	R	NG	36	Imm.	0.16	0.13	0.11	0.23				
35 - Lum & Morton, 1984																			
Spelling Mastery	Misc	Rg.cls.	Class	1 yr.(20-30 m/d)	2nd	N	NG	NE	No	36	Imm.	0.38	0.31		0.45				
36 - Mantzicopoulos et al., 1992																			
Phonetic read/spell	Misc	Rg.cls.	Tutor	50s (1h/wk)	1st	AR	Mid	R	No	112	Imm.	0.53					0.53		
Phonetic read/spell										112	follow up	0.32		0.33	0.3	0.08	0.56		
37 - Marston et al., 1995																			
Direct Instruction	Syn	Rg.cls.	Class	10 wks (45m/d)	gr 1-6	LA	NG	NE	Y/Adj	53	Imm.	0.01						0.01	
38 - Martinussen & Kirby, 1998																			
Successive phonics	Syn	Rg.cls.	SmG	8 wks(40-60m/wk)	K	AR	NG	R	No	26	Imm.	0.62	0.53	0.63	0.68		0.62		
41 - Oakland et al., 1998																			
Orton-Gillingham	Syn	Rg.cls.	SmG	2 yrs (350h)	M=11y	RD	NG	NE	Yes	48	2nd yr tr.	0.54	0.71		0.23	0.62	0.61		

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
44 - Santa & Hoien, 1999																			
RRD-Early Steps	LU	Wh.L.	Tutor	1 yr.(30m/d)	1st	AR	Var	NE	No	49	Imm.	0.76	0.93		0.63	0.73			
RRD-Early Steps										41	follow up	0.86	0.57			0.87	1.15		
47 - Silberberg et al., 1973																			
Lippincott	Syn	Wh.W.	SmG	1 yr.	gr 3	RD	NG	NE	Yes	69	Imm.	0.5	0.7			0.36		0.45	
Orton-Gillingham	Syn	Wh.W.	SmG	1 yr	gr 3	RD	NG	NE	Yes	65	Imm.	0.04	0.31			0.09		-0.29	
Lippincott										62	follow up	0.33	0.37			-0.04		0.66	
Orton-Gillingham										58	follow up	-0.47	-0.19			-0.81		-0.4	
48 - Snider, 1990																			
Direct Instruction	Syn	Basal	SmG	1yr.(60m/d)	1st	N	Mid	NE	No	66	follow up	0.38		0.6	0.44	0.1			
51 - Torgesen et al., 1999																			
Lindamood PA	Syn	Rg.cls.	Tutor	2.5 yrs.(80m/-wk)	K	AR	NG	R	No	65	Imm.	0.33	0.08				0.58		
Embedded	LU	Rg.cls.	Tutor	2.5 yrs.(80m/-wk)	K	AR	NG	R	No	68	Imm.	0.32	0.52				0.12		
Lindamood PA										65	2nd yr tr.	0.75	0.64			0.49	1.13		
Embedded										68	2nd yr tr.	0.28	0.24			0.29	0.31		
Lindamood PA										65	3rd yr tr.	0.67	0.67		0.64	0.36	1.01		
Embedded										68	3rd yr tr.	0.17	0.25		0.1	0.17	0.16		

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
52 - Traweek & Berninger, 1997																			
Direct Instruction	Syn	Wh.L.	Class	1yr.	1st	AR	Low	NE	Y/Adj	38	Imm.	0.07	0.07						
53 - Tunmer & Hoover, 1993																			
RRD-Phonograms	LU	Rg.cls.	Tutor	42 s (30m/d)	1st	AR	NG	NG	NG	64	Imm.	3.71	2.94		1.63		1.49	8.79	
54 - Vandervelden & Siegel, 1997																			
Developmental	Misc	Rg.cis.	SmG	12wks(30--45m/wk)	K	AR	Low	NE	No	29	Imm.	0.47	0.04		1.11		0.57	0.15	
55 - Vickery et al., 1987																			
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	3rd	N	NG	NE	NG	63	Imm.	0.04							0.04
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	4th	N	NG	NE	NG	71	Imm.	0.04							0.04
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	5th	N	NG	NE	NG	74	Imm.	0.61							0.61
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	6th	N	NG	NE	NG	79	Imm.	0.43							0.43
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	3rd	LA	NG	NE	NG	46	Imm.	0.63							0.63
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	4th	LA	NG	NE	NG	47	Imm.	0.19							0.19
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	5th	LA	NG	NE	NG	45	Imm.	-0.2							-0.2
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr.(55 m/d)	6th	LA	NG	NE	NG	41	Imm.	0.13							0.13
57 - Wilson & Norman, 1998																			
Sequential phonics	Syn	Wh.L.	Class	1 yr.	2nd	N	NG	NE	No	54	Imm.	-0.47	-0.33			-0.61			

Appendix G (continued)

	Characteristics of Training				Characteristics of Part.			Features of Design				Effect Sizes on Post-tests							
Author and Year, Treatment	Type of Phonics	Control Group	Tr.unit	Length of Training	Grade/ Age	Reading Ability	SES	Group Assign.	Sig Pre-test Diff	Total N	Time of Post-test	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
59 - Freppon, 1991																			
Sequential phonics	Misc	Wh.L.	Class	1 yr.	1st	N	Mid	NE	Yes	24	Imm.	0						0	
60 - Griffith et al., 1992																			
Traditional basal	Misc	Wh.L.	Class	1 yr.	1st	N	NG	NE	No	24	Imm.	-0.33	-1.11		-0.54	-0.43	0.78		
69 - Umbach et al., 1989																			
Direct Instruction	Syn	Basal	SmG	1 yr.(50 m/d)	1st	AR	Low	R	No	31	Imm.	1.19	1.3			1.08			
72 - Gersten et al., 1988																			
Direct Instruction	Syn	Rg.cls.	Class	4 yrs.	K	AR	Low	NE	No	101	4th yr tr	0.24			0.16	0.28			0.27
Direct Instruction	Syn	Rg.cls.	Class	3 yrs.	1st	AR	Low	NE	No	141	3rd yr tr	0			-0.12	0.11			0.02
74 - Stuart, 1999																			
Jolly Phonics	Syn	Wh.L.	Class	12 wks(60m/-d)	K	AR	Low	NE	Y/Adj	112	Imm.	0.73	0.56		1.11	0.36	0.9		
Jolly Phonics										112	follow up	0.28	0.11		0.5	0.31	-0.03	0.49	
75 - Lovett et al., (in press)																			
Dir. Instruction + Analogy	Com	Misc.	SmG	70h	6-13yr	RD	Var	R	NG	37	Imm.	0.6	0.36	1	0.15	0.27	1.22		
Analogy + Direct Instruction	Com	Misc.	SmG	70h	6-13yr	RD	Var	R	NG	32	Imm.	0.21	0.04	0.55	-0.2	0.12	0.52		
Lovett Direct Instruction	Syn	Misc.	SmG	70h	6-13yr	RD	Var	R	NG	40	Imm.	0.24	0.21	0.36	-0.19,	0.42	0.42		
Lovett Analogy	LU	Misc.	SmG	70h	6-13yr	RD	Var	R	NG	42	Imm.	0.5	0.47	0.75	0.01	0.6	0.66		

Appendix G (continued)

Abbreviations Key

Following is a key to Appendix G.

Word ID = Word Identification
Dec = Decoding
Spell = Spelling
Comp = Comprehension
Nonw = Nonword reading
Oral Read = Oral reading
Gen. Read = Generic reading
Syn = Synthetic;
LU = Larger Units
Misc = Miscellaneous
Com = Combination
Wh.W. = Whole Word
Wh.L. = Whole Language
Rg. Cls. = Regular class
SmG = Small group
yr. = year
m = minutes
m/d = minutes a day

h = hour
s = session(s)
wks = weeks
gr = grade
M = mean
K = Kindergarten
RD = Reading Disabled
AR = At Risk
LA = Low Achievement
NG = Not Given
Var = Varied
Mid = Middle class
R = Random assignment
NE = Non Equivalent groups
Y/Adj = Yes, but means were adjusted for pretest differences
Imm. = Immediate
tr = training
*class was used as the unit of analysis

CHAPTER 3

Fluency

FLUENCY

Executive Summary

Introduction

Fluent readers can read text with speed, accuracy, and proper expression. Fluency depends upon well developed word recognition skills, but such skills do not inevitably lead to fluency. It is generally acknowledged that fluency is a critical component of skilled reading. Nevertheless, it is often neglected in classroom instruction. That neglect has started to give way as research and theory have reconceptualized this aspect of reading, and empirical studies have examined the efficacy of specific approaches to teaching fluency. Here the National Reading Panel (NRP) will provide a summary of the evidence supporting the effectiveness of various instructional approaches that are intended to foster this essential ingredient in successful reading development.

The purpose of this report of the NRP was to review the changing concepts of fluency as an essential aspect of reading, and to consider the effectiveness of two major instructional approaches to fluency development and the readiness of these approaches for wide use by the schools. The first major approach that was analyzed includes procedures that emphasize repeated oral reading practice or guided repeated oral reading practice. These procedures include repeated reading (Samuels, 1979), neurological impress (Heckelman, 1969), radio reading (Greene, 1979), paired reading (Topping, 1987), and a variety of similar techniques aimed at developing fluent reading habits. The second major approach considered here includes all formal efforts to increase the amounts of independent or recreational reading that children engage in, including sustained silent reading programs (Hunt, 1970), the Accelerated Reader (Advantage Learning Systems, 1986), and various incentive programs (i.e., S. Shanahan, Wojciehowski, & Rubik, 1998).

There were a number of reasons why the NRP selected fluency for review and analysis. One is that there is growing concern that children are not achieving fluency in reading. Recently, the National Assessment of Educational Progress conducted a large study of the status of fluency achievement in American education (Pinnell et al., 1995). That study examined the reading fluency of a nationally representative sample of fourth graders, and found 44% of students to be disfluent even with grade-level stories that the students had read under supportive testing conditions. And furthermore, that study found a close relationship between fluency and reading comprehension. Students who are low in fluency may have difficulty getting the meaning of what they read. Given this, it is not surprising that the National Research Council report, Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998), states "Adequate progress in learning to read English (or, any alphabetic language) beyond the initial level depends on sufficient practice in reading to achieve fluency with different texts" (p. 223), and that it recommended, "Because the ability to obtain meaning from print depends so strongly on the development of word recognition accuracy and reading fluency, both the latter should be regularly assessed in the classroom, permitting timely and effective instructional response when difficulty or delay is apparent" (p. 7).

Background

There is common agreement that fluency develops from reading practice. What researchers have not yet agreed upon is what form such practice should take to be most effective. For example, one approach is to have students read passages orally with guidance and feedback. Programs in this category include *repeated reading, neurological impress, paired reading, shared reading,* and *assisted reading,* to note the most popular procedures.

Another, less explicit, but widely used approach, is to encourage students to read extensively on their own or with minimal guidance and feedback. Programs in this category include all efforts to increase the amounts of independent or recreational reading including sustained silent reading (SSR), Drop Everything and Read, Accelerated Reader (AR), and various incentive programs. Often these approaches have no formal name, but take the form of requirements that students engage in unsupervised independent reading at school or home. This report examined the evidence concerning the effectiveness of both guided oral reading procedures and approaches that encourage students to read more.

Methodology

How Was the Analysis of the Research Literature Conducted?

The NRP conducted an extensive and systemic literature review on these two approaches to the development of fluency. Using the methodology and criteria developed for this purpose by the NRP, to reach its conclusions on the effectiveness of each approach, the Panel included only:

1. Studies that were experimental tests of the procedures under examination.
2. Studies that were conducted with students in kindergarten through grade 12.
3. Studies that had appeared in a refereed journal.
4. Studies that had been carried out with English language reading.

Each study which met these criteria was summarized and coded. Where appropriate, the studies were analyzed for their effect sizes, as this allowed the Panel to determine quantitatively the amount of difference such procedures made in children's reading development. Studies that could not be analyzed quantitatively were also examined in order to evaluate the consistency of their findings with those obtained from the quantitative studies.

In its work, the Panel searched two separate databases: PsycINFO and ERIC. The search using PsycINFO identified 1,260 potential articles on instructional PsycINFO approaches to teaching repeated oral reading. This number was deemed too large to search efficiently, so the Panel limited its search to articles that had been published since, and including, 1990. This reduced the number of articles for this topic to 346. A parallel search using ERIC identified 410 potential articles. Removing redundant articles between the two databases resulted in 364 unique articles. Review of each of these article's adherence to the NRP criteria resulted in a total of 77 articles that were coded for possible use in the final analysis.

A similar search process was carried out to identify and locate articles on the effectiveness of encouraging independent silent reading practice. Search of the PsycINFO database identified 478 articles, while the ERIC database identified 325 articles. Removing redundant articles resulted in 603 unique articles on instruction in the various approaches to encouraging independent reading practice. Review of each of the article's adherence to the NRP criteria resulted in the identification of 92 articles. Further careful analysis of these articles according to their adherence to the methodology of the NRP selection procedures resulted in further reduction, with a resulting 14 of which could be used in the meta-analysis to address the Panel's question of whether this instructional approach has proven to be effective in improving reading fluency. Additionally, this analysis was bolstered through a qualitative analysis of 37 other studies that also met these criteria but that could not be used in the meta-analysis for various reasons. These studies were checked for their consistency of findings with those analyzed in the meta-analysis.

As a result of the limitations of the number and quality of studies examining the effectiveness of encouraging independent reading, a meta-analysis was appropriate only in examining the effectiveness of repeated oral reading instructional approaches. In the meta-analysis, the primary statistic used was "effect size," indicating the extent to which performance of the treatment group is greater than performance of the control group. For example, an effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of guided oral reading instruction. In contrast, an effect size of 0 indicates that treatment and control group means were identical and that the treatment had no measurable effect on measured reading performance. In practice, the strength of an effect size can be gauged: a value of 0.20 is considered small; 0.50 is moderate, and 0.80 is large. When available, effect sizes were calculated to determine whether repeated oral reading improved children's accuracy, fluency, and comprehension.

Results and Discussion

What Do the Results of the Analysis of Studies on the Development of Fluency Show?

Are guided repeated oral reading procedures effective in improving reading fluency and overall reading achievement?

The answer was a clear yes. The analysis of guided oral reading procedures led to the conclusion that such procedures had a consistent, and positive impact on word recognition, fluency, and comprehension as measured by a variety of test instruments and at a range of grade levels.

What do results of the meta-analysis of guided oral reading procedures show?

Overall, the study found a weighted effect size average of 0.41, suggesting that guided oral reading has a moderate impact upon reading achievement. Analysis indicated that repeated reading procedures have a clear impact on the reading ability of non-impaired readers through at least grade 4, as well as on students with various kinds of reading problems throughout high school. All approaches were associated with positive effect sizes; however, the sample sizes were generally too small to carry out further analyses comparing one treatment to another within this category.

The interventions demonstrated somewhat differential effects on reading outcomes. The highest impact was on reading accuracy, with a mean effect size of 0.55; the next was on reading fluency, with a mean effect size of 0.44, and the least, but still impressive impact was on reading comprehension, where the effect size was 0.35. In studies where these reading outcome measures were aggregated, the mean effect size was 0.50. These data provide strong support for the supposition that instruction in guided oral reading is effective in improving reading.

Is there evidence that encouraging children to read on their own is effective in increasing reading fluency and overall reading achievement?

The NRP also examined the accumulated research literature on the effects of programs (for example, Sustained Silent Reading and Accelerated Reader) that encourage children to read on their own. The Panel was able to locate relatively few studies on this topic, and these tended to address a narrow range of procedures. The studies examined the impact of encouraging independent reading on overall reading, rather than on reading fluency, per se. Most of these studies failed to find a positive relationship between encouraging reading and either the amount of reading or reading achievement. Furthermore, few of the studies actually monitored the amount of reading students did in the program; therefore, it is unclear whether the interventions led to more reading, or just displaced other reading that students might have done otherwise. Based on the existing evidence, the NRP can only indicate that while encouraging students to read might be beneficial, research has not yet demonstrated this in a clear and convincing manner.

Conclusions

What Conclusions Can Be Drawn From This Analysis of Fluency Development Studies?

Can fluency be encouraged through instructional procedures?

Yes. An extensive review of the literature indicates that classroom practices that encourage repeated oral reading with feedback and guidance leads to meaningful improvements in reading expertise for students—for good readers as well as those who are experiencing difficulties.

Implications for Reading Instruction

Is It Important to Increase Fluency?

Teachers need to know that word recognition accuracy is not the end point of reading instruction. Fluency represents a level of expertise beyond word recognition accuracy, and reading comprehension may be aided by fluency. Skilled readers read words accurately, rapidly and efficiently. Children who do not develop reading fluency, no matter how bright they are, will continue to read slowly and with great effort.

Are These Results Ready for Implementation in the Classroom?

Yes, the NRP found that a range of well-described instructional approaches to encouraging repeated oral reading result in increased reading proficiency. These approaches are well documented and referenced here.

In contrast, the NRP did not find evidence supporting the effectiveness of encouraging independent silent reading as a means of improving reading achievement.

The results of this study indicate that teachers should assess fluency regularly. Both informal as well as standardized assessments of oral reading accuracy, rate and comprehension are available and referenced in the full report.

The demonstrated effectiveness of guided oral reading compared to the lack of demonstrated effectiveness of strategies encouraging independent silent reading suggests the importance of explicit compared to more implicit instructional approaches for improving reading fluency.

Directions for Further Research

The National Reading Panel's extensive review demonstrated good reason to provide instruction encouraging the development of fluency and overall reading proficiency, and indicated which specific approaches the evidence supports as being most effective in increasing fluency. However, this review reveals important gaps in our knowledge. Future research is necessary to address some of these questions.

Research is needed to address the question of the relationship between guided oral reading instruction and the development of fluency. What elements of instructional practice are most responsible for improved fluency? Research is needed to attempt to disentangle the particular contributions of components of guided reading, such as oral reading, guidance, repetition, and text factors. And it is important to know for which children, at what level of reading ability and in what setting and by whom (teachers, classroom aides, peers, parents) and for how long do different approaches to guided oral reading work best?

Research is needed over longer time spans to provide information about the emergence of fluency and its relationship to specific instructional practices. And where along the development of reading are what specific approaches to encouraging fluency most effective?

Research is needed to study in more analytic and rigorous ways, the impact of independent reading on a range of reading outcomes. Since encouraging independent reading is so intuitively appealing and so frequently recommended, it is critical to clarify in a more definitive way the relationship between programs that encourage independent reading and reading development. There is a clear need for rigorous experimental research on the impact of programs that encourage reading on different populations of students at varying ages and reading levels using several different reading outcomes, including amount of reading and specific components of reading achievement, and where the amount of independent reading is carefully monitored.

FLUENCY

Report

The purpose of this report of the NRP is to review the changing concepts of fluency as an essential aspect of reading and to consider the effectiveness of two major instructional approaches to fluency development and the readiness of these approaches for wide use by the schools: first, procedures that emphasize repeated oral reading practice or guided repeated oral reading practice; and second, all formal efforts to increase the amounts of independent or recreational reading that children engage in, including sustained silent reading programs. Because of the fundamental differences in these two approaches, and because of the differing amounts and nature of the articles in these two areas, the Panel was able to perform meta-analysis only on studies relevant to the first topic, repeated oral or guided reading. There were too few experimental studies of the variety of approaches to silent reading for such an analysis; therefore, the Panel performed a more informal analysis of these studies, but felt that some discussion of the studies was nonetheless important.

As a result of these different types of analyses, this report is organized in a slightly different way from the other subreports by the Panel. First, an overall introduction addresses the importance of the development of fluency in reading and provides background for two subsections. From that point, the report is organized in two major sections, with individual methods, results and discussion, implications for reading instruction and directions for future research. Finally, the Panel offers overall conclusions on extant research addressing reading fluency.

Introduction

Fluency, the ability to read a text quickly, accurately, and with proper expression, has been described as the "most neglected" reading skill (Allington, 1983), and with good reason. For much of the 20th century, researchers and practitioners alike assumed that fluency was the immediate result of word recognition proficiency, so efforts were directed towards the development of word recognition, whereas fluency itself was largely ignored. That neglect has started to give way during the past three decades as research and theory have reconceptualized this aspect of reading performance. Research has increasingly turned towards considerations of how instruction and reading experience contribute to fluency development.

The purpose of this report is to review the changing concepts of fluency as an essential aspect of reading and to consider the effectiveness of two major instructional approaches to fluency development and the readiness of these approaches for wide use by the schools. The first major approach that will be analyzed here includes procedures that emphasize repeated oral reading practice or guided repeated oral reading practice. These procedures include repeated reading (Samuels, 1979), neurological impress (Heckelman, 1969), radio reading (Greene, 1979), paired reading (Topping, 1987), and a variety of similar techniques aimed at developing fluent reading habits. The second major approach considered here includes all formal efforts to increase the amounts of independent or recreational reading that children engage in, including sustained silent reading programs (Hunt, 1970), the Accelerated Reader (Advantage Learning Systems, 1986), and various incentive programs (i.e., Shanahan, Wojciehowski, & Rubik, 1998).

Why is fluency important and how well are students doing in achieving fluency? The National Assessment of Educational Progress conducted a large study of the status of fluency achievement in American education (Pinnell et al., 1995). That study examined the reading fluency of a nationally representative sample of 4th graders and found 44% of students to be disfluent even with grade-level stories that the students had read under supportive testing conditions. Moreover, that study found a close relationship between fluency and reading comprehension. Students who are low in fluency may have difficulty getting the meaning of what they read. Given this, it is not surprising that the National Research Council report, Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998), states "Adequate progress in learning to read English (or any alphabetic language) beyond the initial level depends on sufficient practice in reading to achieve fluency with different texts" (p. 223), and that it recommends, "Because the ability to obtain meaning from print depends so strongly on the development of word recognition accuracy and reading fluency, both should be regularly assessed in the classroom, permitting timely and effective instructional response when difficulty or delay is apparent" (p. 7).

Changing Concepts of Fluency

Over the past three decades, our understanding of what is involved in reading fluency has been altered and enlarged. One finds, for example, in the 1974 LaBerge and Samuels' article on automatic information processing in reading, an emphasis on word recognition. This same focus persists in the The Literacy Dictionary *definition* (Harris & Hodges, 1995) that states that fluency is "freedom from word identification problems." More recent conceptualizations of fluency, however, have been extended beyond word recognition and may embrace comprehension processes as well (Thurlow & van den Broek, 1997).

In its early conception, it was recognized that fluency requires high-speed word recognition that frees a reader's cognitive resources so that the meaning of a text can be the focus of attention. However, it is now clear that fluency may also include the ability to group words appropriately into meaningful grammatical units for interpretation (Schreiber, 1980, 1987). Fluency requires the rapid use of punctuation and the determination of where to place emphasis or where to pause to make sense of a text. Readers must carry out these aspects of interpretation rapidly—and usually without conscious attention. Thus, fluency helps enable reading comprehension by freeing cognitive resources for interpretation, but it is also implicated in the process of comprehension as it necessarily includes preliminary interpretive steps.

Early Research on Expertise and Fluency

Recognition of the importance of automatic processes and reading fluency is not new to psychology or education. During the last century, and certainly in the last 30 years, there has been interest in skills acquisition and expertise. Many early investigations of expertise focused on perceptual-motor skills. For example, the Principles of Psychology (James, 1890) explained the importance of practice and repetition in the development of the skills that enabled someone to perform complex acts with ease, and the Bryan and Harter (1899) studies described how telegraph operators learned to send and receive Morse code accurately in larger and larger units.

Not all research was carried out during this early period addressed psychomotor behavior, however. Huey's (1905) book on the reading process became a classic in the field in part because it summarized the research findings of the 1800s on word recognition and eye movements during reading and in part because it was the harbinger for what would later develop into the cognitive psychology paradigm. In that work, Huey made the following perceptive observation about the development of fluency:

> Perceiving being an act, it is, like all other things that we do, performed more easily with each repetition of the act. To perceive an entirely new word or other combination of strokes requires considerable time, close attention, and is likely to be imperfectly done, just as when we attempt some new combination of movements, some new trick in the gymnasium or new "serve" at tennis. In either case, repetition progressively frees the mind from attention to details, makes facile the total act, shortens the time, and reduces the extent to which consciousness must concern itself with the process (p. 104).

From about 1910 until the middle of the 1950s, during what we now designate as the period of "Behaviorism," little research was done on automaticity or reading fluency. Researchers who worked within psychology's behavioral paradigm tended to shy away from research on reading as a psychological process. But, by the 1970s, the pendulum had moved away from behaviorism and back to studies of "inside-the-head" phenomena such as problemsolving and reading. As a result, cognitive psychologists of the period again considered issues such as letter recognition (Posner & Snyder, 1975) and lexical access (Neely, 1977).

It was during this period that linguists attempted to describe the reading process. Fries (1962), for example, discussed the importance of mapping spoken language onto print within reading. According to Fries, to be considered a fluent reader, a person has to do this language mapping rapidly and easily. Soon after, LaBerge and Samuels (1974) published their general theory of automatic information processing in reading in which they explained why automaticity in word recognition was an important prerequisite to skilled reading comprehension. This insight was echoed and expanded in later work.

By this point, theoreticians began to wonder about how fluency skills develop. Stanovich (1990), for example, was critical of assumptions regarding cognitive resource limitations, and Logan's (1997) instance theory explained how a single exposure to a word could leave a sufficient memory trace to allow it to be recognized automatically in the future.

Defining Automaticity and Fluency

There has been a high degree of overlap in the use of terms such as "automaticity" and "fluency." Most scholars treat automaticity as the more general term that embraces a wide variety of behaviors, ranging from motor skills such as driving and typing to cognitive skills such as reading. Some would prefer to reserve the term "fluency" for reading or other language phenomena. This distinction, however, is not universally recognized. For example, The Literacy Dictionary (Harris & Hodges, 1995) defines "fluency" as "freedom from word identification problems that might hinder comprehension . . ." whereas, in the same source, "automaticity" is defined as "fluent processing of information that requires little effort or attention." In other words, automaticity and fluency are often used synonymously.

Actually, the fundamental idea of automaticity requires much more than that information be processed with little effort or attention. This definition has the advantage of simplicity, but it suffers from the fact that it includes within its scope acts that result from innate forces. For example, many behaviors would fall within this definition of automaticity—such as the avoidance of a steep dropoff by newborn mountain goats or the eye blinking and avoidance behaviors exhibited by 3-week-old infants at the rapid approach of a looming object—even though these are not highly skilled expert behaviors. A proper definition of automaticity would rule out behaviors that can be carried out without much previous experience. Automaticity involves the processing of complex information that ordinarily requires long periods of training before the behavior can be executed with little effort or attention. This definition would include various reading behaviors or processes because it is clear that it takes a considerable period of time and substantial practice before even the fastest learners can be considered to be fluent readers.

Furthermore, researchers have generated property lists that can be used to distinguish automatic from non-automatic processes. According to Logan (1997), "The general strategy was to find a list of properties that could be used to define and diagnose automaticity, so that processes, tasks, or performances that possessed those properties could be designated 'automatic,' and processes, tasks, and performances that did not possess them could be designated 'non-automatic' " (p. 124).

One such list described three general properties essential to automaticity (Posner & Snyder, 1975), indicating that the behavior be carried out without immediate intention, without conscious awareness, and without interfering with other process that are occurring at the same time. Shiffrin and Schneider (1977) augmented this list to include two additional properties. They claim that automatic processes are acquired gradually as the result of extended practice and that once activated these processes continue to completion because they are difficult to suppress. The importance of practice in the development of automaticity is also evident in Ackerman's (1987) description:

> Automatic processes are characterized as fast, effortless (from a standpoint of allocation of cognitive resources), and unitized (or proceduralized) such that they may not be easily altered by a subject's conscious control, and they may allow for parallel operation with other information processing within and between tasks. . . *These processes may be developed only through extensive practice under consistent conditions, which are typical of many skill acquisition situations* [p. 4, emphasis added].

Logan (1997) applies the automaticity construct to reading directly by highlighting the role of speed, effortlessness, autonomy (i.e., ability to be completed without intention or deliberation), and lack of consciousness or awareness, although he fails to emphasize the importance of practice or repetition within his description. However, Logan emphasizes one more essential dimension of automaticity in reading that makes his contribution essential to this discussion.

> The property list approach defines automaticity in terms of a list of binary-opposite properties. . . . This view has suggested to some that automatic processes should share all of the properties associated with automaticity (i.e., they should be fast, effortless, autonomous, and unconsciousness) (Logan, 1997).

However, according to Logan, automaticity should be viewed as a continuum rather than a dichotomy. This distinction has important implications for reading.

To show the importance of thinking of fluency as a continuum, consider reading speed as one example. Reading speed at the early stages of instruction tends to be slow and even labored. However, if we examine a student after years of practice, we will typically find that a rapid rate of reading speed has been attained. Was the shift from slow to fast an abrupt one in which the reader was transformed from a nonfluent to a fluent reader, or was this a more gradual change? This question can be answered using data gathered as children practice reading over time. Such data reveal a gradual, continuous improvement in reading speed in which only the beginning and end points could be justifiably characterized as "slow" or "fast." Reading speed, like other aspects of fluency or other automatic behaviors, shows gradual or incremental improvement through practice (Samuels, 1979).

Beyond Accuracy to Automaticity: Why Automatic Decoding Matters

One of the key reasons for the abiding interest in the word recognition process is the consistent finding that development of efficient word recognition skills is associated with improved comprehension (Calfee & Piontkowski, 1981; Herman, 1985; Stanovich, 1985). To understand how efficient word recognition skills can influence other reading processes such as comprehension, word recognition must be fractionated into its component elements such as accuracy of word recognition and the automaticity of word recognition. In the early stage of reading instruction, the beginning reader may be accurate in word recognition but the process is likely to be slow and effortful. With increased practice and repeated exposure to the words in the texts that the student reads, word recognition continues to be accurate but there would be improvements evident in the speed and ease of word recognition as well. Continued reading practice helps make the word recognition process increasingly automatic. In some situations, however, teachers may persist in trying to develop a high degree of word recognition accuracy without commensurate attention to other essential dimensions of fluency (i.e., speed, expression) or may accept recognition accuracy as a sufficient outcome of instruction without any emphasis on true fluency. Although accuracy in word recognition is, indeed, an important reading milestone, accuracy is not enough to ensure fluency—and without fluency, comprehension might be impeded.

Why do problems with reading accuracy, speed, and expression interfere with comprehension? To answer this question, we need to examine the reading process in terms of two basic cognitive tasks. The reader must recognize the printed words (decoding) and construct meaning from the recognized words (comprehension). Both decoding and comprehension require cognitive resources. At any given moment, the amount of cognitive resources available for these two tasks is restricted by the limits of memory. If the word recognition task is difficult, all available cognitive resources may be consumed by the decoding task, leaving little or nothing for use in interpretation. Consequently, for the nonfluent reader, difficulty with word recognition slows down the process and takes up valuable resources that are necessary for comprehension. Reading becomes a slow, labor-intensive process that only fitfully results in understanding.

The reading task for the fluent reader is easier than the one facing the nonfluent reader. After considerable practice, the fluent reader has learned how to recognize the printed words with ease and speed, and few cognitive resources are consumed in the process. In essence, the reader has become automatic at the word recognition task. Because the cognitive demands for word recognition are so small while the word recognition process is occurring, there are sufficient cognitive resources available for grouping the words into syntactic units and for understanding or interpreting the text. The fluent reader is one who can perform multiple tasks—such as word recognition and comprehension—at the same time. The nonfluent reader, on the other hand, can perform only one task at a time. The "multitask functioning" of the fluent reader is made possible by the reduced cognitive demands needed for word recognition and other reading processes, thus freeing cognitive resources for other functions, such as drawing inferences.

Being an "automatic" or "fluent" reader should not be thought of as a stage of development in which all words can be processed quickly and easily. Even highly skilled readers may encounter uncommon, low-frequency words such as oenology, epistrophe, anfractuous, faience, casuistically, and contralesional—words that they cannot recognize automatically but that require some reliance on decoding strategies. Skilled readers usually have several options available for word recognition. They can recognize words automatically or, in cases like these, they can use controlled effortful strategies to decode the word. Unskilled readers, on the other hand, are limited to controlled effortful word recognition.

Research on the eye in the past 2 decades has provided a perspective from which to observe the fluent reading process. These studies take a picture of how the eye moves and what it fixates on during reading. For the most part, readers—no matter how fluent—have to fixate on or look at each word in a text. However, more skilled readers come to fixate on function words (words such as of, the, to, etc.) less often than on content words. It is not so much that fluent readers skip function words as that their facility with such words allows them to see them adequately at the edge of their visual field—while fixating on other words—without having to stop to look at them specifically (Carpenter & Just, 1983; Rayner & Duffy, 1988; Radach & Kempe, 1993). Skilled readers also get better at seeing a word in a single fixation; therefore, they evidence fewer refixations on the same words and fewer short regressions in which they have to come back to look at a word again after they have read other words (Frazier & Rayner, 1982; Kennedy, 1983; Kennedy & Murray, 1987a, 1987b; Murray & Kennedy, 1988). Skilled readers learn to develop a broader perceptual span or word identification span during reading that allows them to take in more information about words in a single fixation (Ikeda & Saida, 1978; McConkie & Rayner, 1975; McConkie & Zola, 1987; Rayner, 1986; Underwood & McConkie, 1985). The placement and overlap of these fixations improve in efficiency as well, allowing fluent readers to integrate the information from each fixation more effectively (McConkie & Zola, 1979; Rayner, McConkie, & Zola, 1980).

Rayner (1998) has summed up the differences in eye movements between good and poor readers:

> There are well-known individual differences in eye movement measures as a function of reading skill: Fast readers make shorter fixations, longer saccades [the jump of the eye from one fixation to another], and fewer regressions than slow readers (Everatt, Bradshaw, & Hibbard, 1998; Everatt & Underwood, 1994; Rayner, 1978b; Underwood, Hubbard, & Wilkinson, 1990) . . . . In characterizing the eye movement patterns of dyslexic readers, Olson, Kliegl, Davidson, & Foltz (1985) categorized such readers as *plodders* and *explorers*; plodders made relatively short forward saccades, and more regressions, whereas explorers showed more frequent word skipping, longer forward saccades, and more regressions (p. 392).

Indicators of Fluent Reading

A number of informal procedures can be used in the classroom to assess fluency. Informal reading inventories (Johnson, Kress, & Pikulski, 1987), miscue analysis (Goodman & Burke, 1972), pausing indices (Pinnell et al., 1995), running records (Clay, 1972), and reading speed calculations (Hasboruck & Tindal, 1992). All these assessment procedures require oral reading of text, and all can be used to provide an adequate index of fluency.

For example, informal reading inventories (IRI) require students to read grade-level passages aloud and silently. The teacher determines a reading level by calculating the proportion of words read accurately in the passage. To ensure that students do not focus solely on fluency—at the expense of comprehension—the student is expected to summarize or answer questions about the text.

The Gray Oral Reading Test–3 (GORT–3) (Wiederholt & Bryant, 1992) is a standardized measure requiring oral reading and providing scoring for reading accuracy, rate, and passage comprehension. In addition, Wagner, Torgesen, and Rashotte (1999) have recently published a standardized measure of word reading efficiency that tests the speeded reading of single words.

The National Assessment of Educational Progress fluency study noted earlier (Pinnell et al., 1995) calculated speed and accuracy but performed most analyses on the basis of a four-point pausing scale. This scale provided a description of four levels of pausing efficiency with one point assigned to readings that were primarily word by word with no attention to the author's meaning, to four points for readings that attended to comprehension and that paused only at the boundaries of meaningful phrases and clauses.

Fluency and Practice

How does one become so fluent in reading that words are recognized accurately, quickly, and with ease and so that a text sounds like spoken language when read aloud? The conventional wisdom is that it is only through extended practice in which large quantities of material are read that the student develops fluency skills that go beyond accuracy of recognition to automaticity of recognition (Allington, 1977, 1984; Snow, Burns, & Griffin, 1998). But how accurate is conventional wisdom? One might assume that with all the research that has been done on factors that produce superior readers, that there would be solid experimental evidence showing a causal connection between input variables such as time spent reading or the amount read and reading outcomes such as fluency.

What is surprising is that most of the evidence linking up input variables such as amount read and output variables such as reading ability is correlational. For example, in a longitudinal study of 54 children, Juel (1988) estimated that 1st grade children with good word recognition skills were exposed to about twice as many words in basal text as children with poor word recognition skills. Biemiller (1977-1978) also reported similar differences in print exposure among readers with different levels of reading ability, and Taylor and her colleagues (Taylor et al., 1999) found that high-achieving primary classes allotted more time for independent reading.

There is ample evidence that one of the major differences between poor and good readers is the difference in the quantity of total time they spend reading. Allington (1977) in his article "If they don't read much, how they ever gonna get good?" found that the students who needed the most practice in reading spent the least amount of time in actual reading. Biemiller (1977-1978) similarly reported substantial ability group differences related to how much reading was done, and Allington (1984) in a sample of first grade students found that as little as 16 words were read in a week by one child in a low-reading group compared to a high of 1,933 words for a child in a high-reading group. Nagy and Anderson (1984) claimed that good readers may read ten times as many words as the poor readers in a given school year. Stanovich (1986), in his article "Matthew effects in reading," suggested that students who start out as poor readers often remain that way. In the Bible chapter on Matthew (Matthew, 25:29), there is the phrase "The rich get richer and the poor get poorer." Stanovich applied this Biblical phrase as a metaphor to reading, claiming poor readers read less than good readers, and he speculated that because of this difference, year after year the gap between the two groups increases. More recent empirical evidence indicates that while poor readers remain poor readers, the gap between the two groups does not increase (Shaywitz et al., 1995).

Although correlational findings may be useful, they also can be deceptive because correlations tell nothing about the direction or sequence of a relationship. That good readers read more could be because reading practice contributes to reading attainment, but it could also be simply that better readers choose to read more because they are good at it. If this is true, then it is reading achievement that stimulates reading practice, not the reverse. Although there is an extensive amount of correlational data linking amount of reading and reading achievement (Cunningham & Stanovich, 1998; Krashen, 1993), such studies do not permit a clear delineation of what is antecedent and what is consequent.

What kinds of practice develop fluency? If fluency were just a word recognition phenomenon, then having students reviewing and rehearsing word lists might make sense. Although there is some benefit to isolated word recognition study of this type, the evidence is that such training is insufficient as it may fail to transfer when the practiced words are presented in a meaningful context (Fleischer, Jenkins, & Pany, 1979). Competent reading requires skills that extend beyond the single-word level to contextual reading, and this skill can best be acquired by practicing reading in which the words are in a meaningful context.

In the sections below, the Panel examines the evidence supporting two major approaches to teaching fluency—first, repeated oral reading and then, silent reading practice.

Repeated Reading and Guided Repeated Oral Reading

Although theories of fluency have emphasized the primacy of practice effects in reading development, most of the evidence has been correlational or ambiguous. Fortunately, several procedures for developing fluency directly through instructional practice have been proposed and evaluated during the past two decades. These procedures typically emphasize repeated reading or guided oral reading practice, including techniques such as repeated reading, neurological impress, radio reading, paired reading, and a variety of other similar procedures. The purpose of each of these procedures is to help students through oral reading practice and guidance to develop fluent reading habits that would allow them to read text more quickly, accurately, and with appropriate expression and understanding.

Historically, most of the instructional attention accorded to oral fluency was developed through round-robin reading, a still widely used approach in which teachers have students take turns reading parts of a text aloud (Opitz & Rasinski, 1998). These procedures have been criticized as boring, anxiety provoking, disruptive of fluency, and wasteful of instructional time, and their use has been found to have little or no relationship to gains in reading achievement (Stallings, 1980). It is evident that with round-robin procedures students receive little actual practice in reading because no child is allowed to read for very long. Such procedures do provide students with some guidance or feedback—although studies suggest that teachers vary greatly in their ability to provide this effectively (Pflaum & Pascarella, 1980). But even when this guidance is of high quality, students rarely have the opportunity to perfect their performance of a passage, as most texts tend to be read only once.

Newer guided repeated oral reading techniques share several key features. First, most of these procedures require students to read and reread a text over and over. This repeated reading usually is done some number of times or until a prespecified level of proficiency has been reached. Second, many of these procedures increase the amount of oral reading practice that is available through the use of one-to-one instruction, tutors, audiotapes, peer guidance, or other means. In round-robin reading, time was severely limited because the teacher was the only one allowed to provide expert guidance; that is not true of the newer procedures. Third, some of the procedures have carefully designed feedback routines for guiding the reader's performance.

The purpose of this section of the review is to provide a research synthesis of empirical studies that have tested the efficacy of repeated reading and other guided oral reading procedures. The Panel's purpose is to determine whether the use of such procedures improves student fluency and whether such improvements are evident in better reading comprehension, how appropriate such procedures would be for regular classroom application, and what additional research is needed.

Repeated and Guided Repeated Oral Reading: Methodology

Database

The Panel determined that the literature search for a research synthesis must be conducted in a systematic, replicable way and that these procedures be described thoroughly. This methodology will allow others to weigh the appropriateness of the procedures for answering the research questions and to check for bias and error.

Consideration of Extant Literature Searches. This search started with the location of two published literature reviews on the impact of repeated reading procedures (Strecker, Roser, & Martinez, 1998: Toward understanding oral reading fluency. Forty-seventh Yearbook of the National Reading Conference (pp. 295-310); Dowhower, 1994: Repeated reading revisited: Research into practice. Reading and Writing Quarterly, 10, 343-358). These literature searches were used in two ways. First, they were examined carefully to identify appropriate terminology that could be used to conduct a thorough electronic search of the literature. Second, the reference lists included in these literature searches were examined for additional, potentially relevant studies on this topic.

Identification of Appropriate Terminology

This search depended on electronic databases, and these require the use of appropriate search terms. In addition to these literature reviews, the NRP examined various published reference sources to help identify terms for use in the search. The Panel used The Literacy Dictionary (Harris & Hodges, 1995); Handbooks of Reading Research I and II (Barr, Kamil, Mosenthal, & Pearson, 1991; Pearson, Barr, Kamil, & Mosenthal, 1984); The Encyclopedia of English Studies and Language Arts (Purves, 1994); and the Handbook of Research on Teaching the English Language Arts (Flood, Jensen, Lapp, & Squire, 1991). These sources were examined for articles on fluency, oral reading, repeated reading, and other relevant topics identified during this analysis and from the previous literature searches.

These efforts led to the identification of terms that described particular instructional approaches, as well as those that focused on specific aspects of reading that supposedly are improved by the application of such procedures. Table 1 provides a list of the 22 search terms that were used in this synthesis.

Table 1

Terms used to search the electronic databases for studies that evaluated the effectiveness of repeated reading and other guided oral reading procedures.

chunking
echo reading
speech pitch
punctuation
reading rate
repeated reading
reading fluency
paired reading
reading speed
automaticity
prosody
parsing
intonation
expression
phrasing
reading accuracy
neurological impress
assisted reading
inflection
verbal fluency
instance theory
oral reading

Electronic Search Strategies

Because of the nature of the topic and the possibility that a single search could miss key information, the Panel elected to examine two separate databases: ERIC and PsycINFO. The Panel searched PsycINFO using the terminology listed in Table 1.

Each of these terms was linked by OR statements, meaning that if any article in that database focused on any of these topics, it would be included in our target pool. The target pool that was identified in this way included 18,763 articles. This number was reduced slightly by limiting the pool to include only English-language articles. Then a separate focus pool was constructed using the terms: reading, reading ability, reading achievement, reading comprehension, reading development, remedial reading, silent reading, reading education, reading materials, reading skills.

These reading topics were linked with each other by OR, again, with the idea of identifying all articles about any aspect of reading in the PsycINFO database. The focus pool included 16,422 English-language articles. This focus pool was then combined with the target pool using AND as the link. This means that the Panel was discarding anything in the target pool that was not clearly linked with reading or reading education. The resulting combination resulted in the identification of 1,260 potential articles.

This number was still deemed too large to search efficiently, so the Panel used number of years as a delimiter. That is, the Panel limited the search to articles in the PsycINFO database that had been published since 1990 (inclusive of 1990). This limit reduced the number of target articles to 346 and printed out abstracts for each of these papers.

Each abstract was read and coded as to whether it should be included in the search for articles. To be included, an article had to meet the following criteria:

1. The study had to examine the impact of repeated reading or some other form of guided oral reading instruction on reading achievement.
2. The study had to focus on reading in English, conducted with children (K-12).
3. The study had to have appeared in a refereed journal.

4. The study had to have been carried out with English-language reading.

If an article was clearly inappropriate in terms of these criteria, it was rejected without search. Rejected articles were designated as (1) nonrefereed, (2) nonresearch, (3) off topic/off sample, or (4) non-English language instruction. Although an abstract might indicate several violations, only one needed to be noted for an article to be rejected. A conservative application of these criteria was used to ensure the inclusion of any article that *might* be tangentially appropriate to our search goals because this would allow us to make sense of articles that could reveal important information about fluency learning. Because of this, analyses of the relationships among various fluency measures, studies of the correlation of fluency and comprehension, or literature searches on related topics were all retained in the pool at this stage. Such articles would not be used for the final analysis of whether guided repeated oral reading procedures are effective, but they were used to help identify relevant studies outside the boundaries of these search procedures. As a result of this screening, the Panel attempted to locate 81 articles for further consideration.

The same basic terminology and search procedures were used in the ERIC system. The search for target pool items was identical to that carried out in PsycINFO. Because ERIC uses a larger collection of reading-relevant terminology, the focus pool was expanded to ensure the widest possible inclusion of reading articles. The focus pool included basal reading, beginning reading, content area reading, critical reading, decoding, directed reading activity, early reading, independent reading, individualized reading, oral reading, reading, reading ability, reading achievement, reading aloud to others, reading comprehension, reading difficulties, reading failure, reading habits, reading improvement, reading instruction, reading material selection, reading materials, reading motivation, reading processes, reading programs, reading rate, reading research, reading skills, reading strategies, recreational reading, remedial reading, silent reading, speed reading, story reading, supplementary reading materials, OR sustained silent reading.

For this search, the target pool included 6,730 potential items. This was reduced to 2,053 items on combination with the focus pool of 39,694 items. This set was further reduced to 840 potential articles by omitting non-English language reports and nonjournal articles. For the sake of consistency, 1990 inclusive was again the cut-off year for the electronic search. This reduced the ERIC search to 410 potential items.

Of these 410 items, a review of the abstracts indicated that only 50 of these had potential value for our purposes. Many of these, however, had already been identified in the PsycINFO search and did not need to be double counted. Thus, the ERIC search resulted in the identification of only 18 additional potential studies or articles.

Location of Articles

As a result of these two searches, the Panel set out to find 99 articles on guided repeated oral reading. Of these, the Panel was able to locate 76 articles, or 77% of the total. Of the articles that could not be located, only 11 met or appeared to meet all of the selection criteria; it was recognized that the other 12 papers did not actually meet the criteria although these papers had some apparent relevance to the topic. Of the 11 papers the abstracts of which suggest that they might have met the criteria, nine abstracts claimed positive and substantial improvements in reading due to the procedures used, one reported no significant difference, and one reported mixed results. It is possible that locating these missing studies could alter the findings of this report. Any alteration, however, would likely strengthen the support for guided oral reading procedures given that the vast majority of these appear to provide evidence on that side of the equation.

Each of the 77 articles that were located was reviewed to determine its relevance to the topic and its adherence to the various selection criteria. Any study that appeared to meet the criteria was then coded for possible use in the final analysis.

Further Identification of Articles

The Panel's search procedures were biased against older studies of these instructional procedures. Only studies that had been published since 1990 were included in the selection procedures up to this point. To expand on that set of studies in an effective manner, the Panel analyzed the reference lists of all studies that were located through the previously described procedures. Even studies that were determined to be in violation of the final selection criteria were analyzed in this way. The literature searches that the NRP used as the starting point for its electronic searches were also examined for relevant references that were not in its search set. This led to the consideration of 133 additional papers, and of these the Panel was able to find 109 or 81%. For the most part, these second-generation papers had been published before 1990. Of these 109 papers, only 21 were found to meet all of the selection criteria. These 21 studies were added to the 77 already identified, and these were designated for further examination and coding.

Analysis

Each of these studies was read and summarized on a six-page coding sheet. Each study was summarized in terms of the following variables: reference, narrative summary, source of citation, states or countries represented in the sample, number of schools included, number of classrooms included, number of participants, number of participants in each group, student ages, student grade levels, reading levels of the participants, community (urban, suburban, rural), socioeconomic status, ethnicity, exceptionality, sample selection criteria, availability of additional reading instruction, amount of attrition per group, how attrition was addressed, study location (classroom, lab, clinic, pullout, other), assignment to groups (random, matching, etc.), sample equivalence, description of each treatment and control condition, nature and difficulty of texts used in treatments, duration of treatments in minutes of training, duration of treatment from beginning to end in days, checks on treatment fidelity, student/teacher ratios, trainer (classroom teacher, researcher, parent, peer, etc.), amount and type of training for trainers, special costs associated with treatment, and pretests and posttests means and standard deviations.

If information was omitted from the original study, it was omitted from the coding. The most serious omissions were evident in the older studies (pre-1994), and no effort was made to locate authors of the original studies to help fill in these gaps. After coding, these data were further summarized within a spreadsheet program (Microsoft Excel) to allow statistical analysis and comparison.

Reliability

A 10% sample (10 articles) was randomly selected for independent re-analysis. The coefficients of agreement ranged from 0.88 to 1.00, with most variables receiving a 1.00. The lowest agreements were evident with student/teacher ratios, trainer identification, and numbers of subjects lost to attrition.

Consistency With the Metholodogy of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. However, the wide variations in methodologies and implementations required the subcommittee to qualify its use of the NRP Criteria for Evaluating Single Studies, Multiple Studies, and Reviews of Existing Studies. These departures from the stated NRP criteria are described below.

Coding these variables made it clear that the studies that were being examined represented dramatically different conceptualizations of the problem. As a result, the NRP divided articles into four sets. One set of 14 articles, Immediate Effects Articles, examined the immediate impact of repeated reading and guided oral reading on a reading performance with no effort to measure transfer to other reading (see Appendix A). To be placed in this set, a study had to examine how reading performance changed with feedback or repetition but with no transfer measure to other passages. These studies are valuable because they examine changes to reading behavior that could contribute to a more general change in reading ability although they do not attempt to measure that change directly.

The second set of articles, Group Experiments, attempted to evaluate the impact of repeated reading and other guided oral reading procedures on the reading abilities of students in grades K to 12 (see Appendix B). To be included in this group, a study had to meet the following criteria:

1. Study had pretest and posttest measures of reading, separate from the material used for training.
2. Study had a treatment group that received some form of guided repeated oral reading training and a comparison group that did not receive such training.

There were 16 articles in this set. These studies could be directly evaluated through meta-analysis to test the claim that guided repeated oral reading procedures improve reading ability.

The third set of articles, Single Subject Studies, used multiple baseline single-subject designs to examine the impact of repeated reading and other guided oral reading procedures on the reading abilities of students in grades K through 12 (see Appendix C). These studies had to have some measure of reading transfer. These studies could be used to directly evaluate the claim that guided oral reading procedures improve reading ability, but they were not used in the meta-analysis. Data from these studies were used to confirm or contradict the meta-analysis results.

The fourth set of studies, Methods Comparisons, compared different methods for doing repeated reading or guided repeated oral reading but did not have a true control group (see Appendix D). These studies were based on the assumption that guided repeated oral reading procedures improve reading ability, and they were usually attempting to discern which methods work best. The lack of control group meant that these studies could not be used to evaluate the claim of whether guided repeated oral reading improves reading ability, but these studies could help guide any further analysis or help determine the applicability of such methods to regular classrooms. There were eight of these studies.

Repeated and Guided Repeated Oral Reading: Results and Discussion

Immediate Effects Articles

There were 14 studies found that dealt with the immediate impact of different programs of repetition and feedback during oral reading on the reading performance of a specific passage or article. It is important to note that these studies did not fail to find transfer effects for these procedures, only that these studies did not attempt to measure such transfer. These studies typically measured some aspects of fluency or comprehension with a particular passage and then monitored changes in this performance from one reading to another. Not surprisingly, all 14 studies reported demonstrable improvements from a first passage reading to a final passage reading with whatever measures were used.

Nine of these studies considered the impact of repeated reading (Faulkner & Levy, 1999; Levy, Nicholls, & Kohen, 1993; Neill, 1979; O'Shea, Sindelar, & O'Shea, 1985; Rasinski, 1990; Sindlar, Monda, & O'Shea, 1990; Stoddard, Valcante, Sindlar, O'Shea, & Algozzine, 1993; Turpie & Parratore, 1995; VanWagenen, Williams, & McLaughlin, 1994), although in other studies, repeated reading was combined with other procedures such as a particular type of oral reading feedback (Reitsma, 1988) or phrasing support for the reader (Taylor, Wade & Yekovitch, 1985). Repeated reading studies either required a set number of repetitions (as few as one and as many as seven) or required students to practice repetition for some amount of time or until some fluency criteria were reached. Other studies had students practicing oral reading while listening to the text being read simultaneously (Bon, Boksebeld, Freide, & van den Hurk, 1991; Rasinski, 1990; Smith, 1979), previewing a text through listening (Reitsma, 1988; Rose & Beatty, 1986), or receiving particular types of feedback during oral reading (Anderson, Wilkinson, & Mason, 1991; Pany & McCoy, 1988).

All these interventions saw clear improvement, although some conditions were better than others. For example, repeated reading with phrasing support seemed to be no better than repeated reading alone in a study of 45 good- and poor-reading 5th graders (Taylor, Wade, & Yekovich, 1985), whereas repeated reading with feedback or guidance (Pany & McCoy, 1988) was superior to repeated reading alone with 3rd graders.

These studies in their totality examined the reading of 752 subjects ranging from 1st grade through college. Four of these studies used normal populations, two compared the performances of good and poor readers, and the rest dealt with students who were somewhat below grade level, substantially behind grade level, or designated as learning disabled. The studies found clear improvements across multiple readings regardless of students' reading levels or age levels although greater gains were sometimes attributed to poor readers. Given the lack of transfer measures in this study, the greater gains for low readers could be an artifact of the design because these readers' initial performances would be relatively more deficient and would therefore be most amenable to improvement.

What inferences can be made from this set of studies? It certainly cannot infer that repeated reading or other guided repeated oral reading procedures would be effective in raising reading achievement on the basis of these studies alone. However, the clear improvements in reading rate, accuracy, and comprehension found for a wide range of readers under a wide range of conditions suggest the possibility that such procedures could have transfer effects worth examining.

Group Experimental Studies: Meta-Analysis

Sixteen studies met the criteria for inclusion in the meta-analysis; these studies met the NRP review methodology. Each of these studies had pre- and post-tests that allowed for an analysis of the improvement or lack of improvement in reading and treatment and control groups that would allow the changes in outcomes to be attributed to the instructional procedures of interest. Of the 16 studies, 2 did not provide sufficient information to allow inclusion in the meta-analysis (Labbo & Teale, 1990; Lorenz & Vockell, 1979) although the findings of these studies will be considered in this section and their data will be included in calculations wherever relevant and possible. The Lorenz and Vockell study found no differences because of the treatments; however, the Labbo and Teale study found clear improvement as a result of repeated reading.

Although these studies were meta-analyzed, this analysis does not go very far. That is, the NRP did not attempt to evaluate all possible comparisons. Such thorough analysis can be informative for future research, but given the national scope of this effort and the potential significance of these determinations, the NRP decided to consider only questions that could be answered with a high degree of certainty (i.e., those that could be answered using all or most of these data). The studies in this set were conducted from 1970 to 1996, and most were carried out in the 1990s.

Calculation of Effect Sizes

Effect sizes were calculated for each relevant comparison. These effect sizes used either the *d index* (Cooper, 1998, p. 128) or the *d index* calculated from the F tests (Cooper, 1998, p. 129). When there were multiple experimental groups in a study, effect sizes were calculated for each guided oral reading group compared with a control group, so if a study had two experimental groups and one control group, there would be two effect sizes for each measure for that study. However, if one of these experimental interventions was not a form of guided repeated oral reading, no effect size would be calculated for that comparison, and those subjects would be dropped from the analysis. Even with these omissions, because most studies included multiple outcomes, 99 effect sizes were calculated for direct comparisons of experimental and control group performance. When multiple-effect-size statistics were calculated for a single study, the mean of effect sizes for that study was calculated to determine a study effect size.

Were Effect Sizes Greater Than Zero?

In all but two of the studies, comparisons resulted in significant differences for the guided repeated oral reading groups over the control groups. Lorenz and Vockell (1979) found no benefit of these procedures for LD students after 13 weeks of neurological impress training with either reading comprehension or vocabulary. The other study that did not result in a positive outcome (Mathes & Fuchs, 1993) compared peer-mediated repeated reading with both peer-mediated silent reading and a control group. There were no significant differences between these treatments with LD students in a special education setting. All other comparisons significantly favored the guided repeated oral reading groups.

Great variance was evident in these study effect sizes; they ranged from as low as 0.05 (almost no effect) to as high as 1.48 (a substantial effect). The average of these study effect sizes was 0.48. However, these studies reported data on as few as 12 subjects and as many as 78. This means that the small studies would have as large an impact on this average as the largest studies. A weighted average is probably more accurate in this case, and it results in a study effect size average of 0.41. The largest effect sizes were obtained with some of the smaller samples, but this is probably an effect of the treatment features of these studies rather than an artifact of sample size. The smaller studies were less likely to use peer tutors; that is the students in the small studies received guidance and feedback from adults (teachers or researchers) rather than from other kids. These effect sizes, weighted or not, suggest that guided oral reading procedures have a moderate impact on the reading achievement of the types of students who participated in these studies.

Characteristics of Students

These 16 studies included data from 752 elementary and secondary education students. The data were drawn from students from six U.S. states and two other countries. The students attended 47 different schools (one study did not report the number of schools so this is an underestimate) and 98 classrooms (again, an underestimate because five studies, including some with relatively large sample sizes, did not provide this information). Not all were included in the analyses, however. As has been noted, two studies provided clear experimental evidence concerning the efficacy of the procedures but failed to include sufficient information for effect size calculation. These studies reported data on 74 subjects, and they were not included in effect size calculations. Also, given that not all comparisons within each study were relevant to our research questions, the Panel dropped from its analysis the data from an additional 73 subjects. Thus, the meta-analysis is based on data from 605 students.

The students in these studies ranged from grade 2 through grade 9. The studies that focused on average reading level samples or normal classroom populations focused on students in grades 2 through 4, while studies of poor readers included students from grades 2 through 9, with most of these drawn from the upper elementary grades. These studies as a collection have not provided sufficient data to allow for a sound analysis of the relative impact of repeated reading procedures on students at different grade levels. It is evident from the studies included in this set that repeated reading procedures have a clear impact on the reading ability of nonimpaired readers at least through grade 4, as well as on students with various kinds of reading problems throughout high school. Future research needs to determine at what point such instruction is no longer beneficial to normal readers.

Eleven of these studies (including the two not used in the meta-analysis) focused on poor readers, whereas only five studied average classrooms. The sample sizes of these studies differed so much, however, that the disparity between numbers of average and poor readers was not as great as this suggests. These 16 studies included 398 students who were selected as poor readers (although data on only 324 of them were used in the meta-analysis) and 281 good readers.

The average effect sizes for these two groups of studies (those examining low-level readers and those that considered average readers) were highly similar and close to the overall average (0.49 for the nine low-level reader studies and 0.47 for the five average-reader studies). When weighted by sample sizes, the average effect sizes diverged more but, surprisingly, the nonimpaired reader studies showed the superior outcomes (0.50 versus 0.33). This is probably attributable, at least in part, to the longer time evident in the nonimpaired reader studies (an average of 24 to 25 hours in nonimpaired reader studies but only about 18 to 19 hours in the poor-reader studies).

Although some of the studies speculated that poor students might benefit more from these procedures, fluency is developmental and students must continue to meet the challenge of increasingly more difficult text as they develop as readers. It is possible, as Faulkner and Levy (1999) have shown, that good and poor students benefit from different aspects of this treatment, with poor readers learning more about the words and good readers developing a stronger command of the prosody of the passages. All of these studies tried to assign students to materials considered to be of appropriate levels of difficulty for the particular students, and this masks or complicates the true meaning of the performance disparity for good and poor readers.

Properties of Instructional Approach

Many different instructional procedures were examined in these studies, so many that it is impossible to determine the best of the few studies. No method was used so often that a reliable estimate of effect size would be possible. Also, variations across studies are subtle in terms of material selection and amount and type of repetition and feedback. Some treatments were delivered by teachers or researchers, some by parents, some by other students, and some by the students themselves with computers or tape recorders. The treatments went under names such as neurological impress, repeated reading, peer tutoring, shared reading, assisted reading, and oral recitation method. All were associated with positive effect sizes. Some might be better, or better in particular circumstances, but the sample sizes associated with any of these associated treatments were too small to allow for a meaningful partialing of variance. Given what is known, all of these procedures seem to have a reasonably high likelihood of success.

Outcome Measures

These studies used a range of outcome measures, including tests of word knowledge, comprehension, and fluency, as well as combinations of these as overall scores derived from standardized reading measures. Some studies had multiple comprehension or fluency measures as well. The Panel attempted to determine whether these guided procedures had a greater impact on some aspects of reading than on others. These studies made 99 different comparisons that were relevant to the analyses. Only one pooled effect size per study per category (word recognition, fluency, comprehension, total score) was drawn from each study, and each of these was weighted by the numbers of subjects whose data were represented in each.

Across these studies, considering all sample comparisons and all measures, there were 49 different comparisons that used some form of comprehension test as an outcome measure. They included standardized tests of reading comprehension in which students read passages and answered multiple choice questions, as well as informal measures such as questions and passages, retellings, and maze tests. The mean weight effect size for these 49 comparisons drawn from 12 separate studies was 0.35.

There were 35 comparisons that used some fluency measure as an outcome. They included standardized tests of reading rate and accuracy, as well as informal measures of these using instruments such as informal reading inventories. The mean weighted effect size for these 35 comparisons drawn from 10 different studies was 0.44.

There were 11 comparisons that used some measure of word recognition. They included standardized tests of word knowledge as well as informal measures that examined students' ability to read particular words or word lists. The mean effect size for these 11 comparisons drawn from eight different studies was 0.55.

Finally, four of the comparisons considered aggregate or full-scale reading scores (these tended to be combinations of the other measures noted above) and included both full-scale scores from standardized tests of reading and reading-level scores derived from informal reading inventories. The average effect size for these four aggregate comparisons from four different studies was 0.50.

Implications for Reading Instruction

As expected, the biggest effect of these procedures was on word recognition and fluency measures, with the smallest effects evident in reading comprehension. It appears that oral reading practice and feedback or guidance is most likely to influence measures that assess word knowledge, reading speed, and oral accuracy. Nevertheless, the impact of these procedures on comprehension (and on total reading scores) is not inconsiderable, and in several comparisons it was actually quite high. These changes in comprehension might take place simultaneously, with the improvements in word recognition and fluency mediating the improvements in comprehension, or there could be a hierarchical order to this, as Faulkner and Levy (1999) have speculated, with the lowest level readers improving in word recognition and the highest ones in comprehension.

Studies Using Single-Subject Designs

Twelve additional studies reported experiments that used single-subject designs. See Appendix C for a list of these studies. The single-subject studies, because of their designs, were not combined in the meta-analysis, although the data were examined to evaluate the conclusions drawn from the meta-analysis. These studies focused on the reading of small groups of students, as few as 2 and as many as 13 (an average of 4 to 5). All these studies addressed the learning needs of elementary grade students with learning problems (i.e., special education, learning disabilities, autism, disfluent readers, readers substantially below grade level). All these studies provided some kind of one-to-one tutoring to students (sometimes parent or peer tutoring) or repeated reading work with tape recorders, for varying lengths of time (as little as 4 weeks and as long as 1.5 years, with most treatments lasting fewer than 10 weeks).

With one exception (Law & Kratochwill, 1993), all these studies found clear and substantial improvements in reading accuracy, speed, or comprehension. The best of these studies calculated a clear reading performance baseline over several days. Then they intervened with repeated reading, oral reading feedback, or reading-while-listening treatments and monitored student growth with new materials during the treatment and with standardized tests at the conclusion. For example, Blum and colleagues (1995) found that the introduction of repeated reading with tape recorders led to marked improvements in student reading performance; that when the training ended, the students maintained their gains; but when the intervention ended, the accelerating improvement ceased. Another example of a well-designed, single-subject study was reported by Kamps and her colleagues (Kamps, Barbetta, Leonard, & Delquadri, 1994).

The one study that found no effects resulting from paired reading of students with parents also found no improvements in word accuracy or reading speed after 6 weeks of treatment. This study had an especially weak design (failed to calculate a stable baseline in student reading performance and did not check on fidelity of treatment). In any event, no gains were found in this study of lst through 3rd grade students.

The pattern of findings for these studies is almost identical to what was reported in the meta-analysis. Most, but not all, of the studies reported clear improvements. The changes described here were a bit larger in magnitude, but all but one of these studies were conducted with a one-to-one teacher-student ratio and all were carried out with low-level—sometimes very low-level—readers, and either of these factors could magnify the effect. Again, the conclusion is that repeated reading and other related oral reading procedures have clear value for improving reading ability.

Methods Comparisons

Nine additional experiments were located that dealt with repeated reading and other guided repeated oral reading procedures. None of these studies used a true control group, however, so it is not clear whether these gains were greater than expected in the amounts of time studied. These studies provided comparisons of the efficacy of various oral reading procedures or were meant as feasibility studies to evaluate the classroom readiness of the procedures.

There were not enough comparisons of guided repeated oral reading procedures to allow for a systematic determination of best procedures. For the most part, the comparisons that were done resulted in no differences. In other words, each of the procedures examined did about as well as the others. Some of the comparisons that were made included repeated reading with and without feedback (Dowhower, 1987), guided repeated reading and assisted nonrepetitive reading (Homan, Lesius, & Hite, 1993), and various peer or parent tutoring procedures in which students read aloud together or read to their parents (Lindsay, Evans, & Jones, 1985; Winter, 1986, 1988). The lack of clear differences among procedures is consistent with the findings of the meta-analysis and again suggests the robustness of these procedures for stimulating reading improvement.

One exception to the no-differences finding, which should be noted, was reported by Rashotte and Torgeson (1985). They did not vary the procedures, but tried out passages that either shared or did not share lots of words with the outcome measures. They found clear gains after 3 weeks for the passages with shared words but not for those without. This suggests that, at least for very poor readers, the first thing that is probably learned from repeated reading is the words (Faulkner & Levy, 1999) and that this growth might be facilitated by using passages that share lots of vocabulary.

Only one study was found that directly evaluated the feasibility of these procedures for use in regular school settings, though several of the studies already noted have done just that. Dixon-Krauss (1995) conducted a feasibility study of partner reading with 24 1st and 2nd graders in regular classrooms. The program proved to be manageable for the regular classroom teachers, and the students were positive about the activity. What was so notable about this study was that it focused on the teacher's abilities to use these procedures on a targeted basis with struggling readers, rather than with whole classes. The findings from this study are consistent with the findings of the other studies that considered classroom effects, including Rasinski's (1990), which had regular classroom teachers applying such procedures on a classwide basis for almost an entire school year. Several other studies showed that regular teachers, with little or no extra training, could successfully use these procedures (for instance, Conte & Humphrey, 1989; Labbo & Teale, 1990; Reutzel & Hollingsworth, 1993; and Shany & Biemiller, 1995). There were also several special education studies in which students provided peer tutoring to their classmates under the direction of their teachers (Mathes & Fuchs, 1993; Simmons et al., 1994; Simmons et al., 1995). Teachers, parents, or peer tutors at most were provided 1 to 4 hours of training, and usually the procedures did not require special materials (though some interventions used tape recorders or elaborate computerized tutoring).

Implications for Reading Instruction

Increasingly, teacher educators and educational researchers and theorists have called for more attention to direct instruction in fluency. Various procedures have been proposed for teaching students to read quickly, accurately, and with proper expression, though it is evident that this remains a serious weakness among many schoolchildren.

A very thorough search for studies that evaluated the efficacy of various guided repeated oral reading procedures was made. Those studies provide a persuasive case that repeated reading and other procedures that have students reading passages orally multiple times while receiving guidance or feedback from peers, parents, or teachers are effective in improving a variety of reading skills. It is also clear that these procedures are not particularly difficult to use; nor do they require lots of special equipment or materials, although it is uncertain how widely used they are at this time. These procedures help improve students' reading ability, at least through grade 5, and they help improve the reading of students with learning problems much later than this.

Repeated and Guided Repeated Oral Reading: Directions for Further Research

There is a need for more research on these issues. Clearly there is a need for longitudinal research that examines the impact of these procedures on the reading development of normal readers at different points along the continuum. The methods used should be characterized not by labels such as repeated reading, but by treatment descriptions that are explicit with regard to how much rereading there is, the nature and timing of the feedback, and the level of difficulty of the materials. Some effort should be made to document the changes that take place in student reading and knowledge during the intervention rather than just at the end.

Longitudinal studies of the impact of these procedures on nonimpaired readers could clarify how long the benefits can be maintained. It would be especially useful if these were examined under various conditions in terms of passage difficulties and feedback procedures. However, given the clear and substantial improvements produced by a wide range of reading procedures, the Panel thinks it advisable that teachers include such activities in their regular instructional routines at least during the elementary grades, and certainly with struggling readers.

One word of caution can be drawn from a short-term study (Anderson, Wilkinson, & Mason, 1991) that found that too much attention to fluency issues within a reading lesson could detract from reading comprehension. It should be noted that in all of these studies, the fluency work was only part of the instruction that students received. In most cases, the fluency work was relatively brief (15 to 30 minutes per lesson), and students who received these lessons were still engaged in other reading activities including comprehension instruction. Guided repeated oral reading and repeated reading provide students with practice that substantially improves word recognition, fluency, and—to a lesser extent—reading comprehension. They appear to do so, however, in the context of an overall reading program, not as stand-alone interventions.

Encouraging Students to Read More

The NRP focused on another widely recommended approach to developing fluent readers—encouraging children to read a lot. Despite all of the controversy about reading instruction, there has been widespread agreement about the value and efficacy of reading practice in developing better readers. The importance of reading as an avenue to improved reading has been stressed by theorists, researchers, and practitioners alike, no matter what their perspectives. There are few ideas more widely accepted than that reading is learned through reading.

And why not? The theories of practice that have already been discussed do not differentiate much between different forms of practice, and so it is unclear why lots of reading would not contribute to improvement. It is possible that oral reading and silent reading operate differently in this regard, but theories of learning to read really do not make much of an issue of this distinction, and theories of practice generally do not stress such differences either. There seems little reason to reject the idea that lots of silent reading would provide students with valuable practice that would enhance fluency and, ultimately, comprehension. Nevertheless, the correlational evidence is overwhelming. There are literally hundreds of studies that find that the best readers read the most and that poor readers read the least; they include the National Assessment for Educational Progress, which has found such relationships with both elementary- and secondary-age students (Donahue et al., 1999). It appears—from the correlations—that the more that you read, the better your vocabulary, your knowledge of the world, your ability to read, and so on.

As a result of such widespread agreement and such clear evidence, books and journals for teachers emphasize ways that teachers can encourage voluntary reading. Several procedures for stimulating students to read more (SSR, DEAR, Million Minutes, etc.) are in the reading education literature and are used with great frequency in the schools. Corporate incentive plans have been widely used to reward students for more reading (e.g., Pizza Hut's Book It), and various programs and materials are available commercially (e.g., Accelerated Reader) that have the purpose of stimulating greater amounts of reading.

There could be a problem with this widespread belief, however. These data are correlational and correlations do not imply causation. That is, it could be that if you read more, you will become a better reader; but it also seems possible that better readers simply choose to read more. So which is it? Well, it is impossible to know from correlational studies alone. For this reason, the NRP chose to examine what effect encouraging students to read would have on student reading achievement. Even if more reading is beneficial, it is possible that programs designed to stimulate greater amounts of reading would fail to have this effect.

The Panel's purpose here is to provide a research synthesis of empirical studies that have tested the efficacy of encouraging reading in terms of its impact on improving reading achievement. The Panel hopes to determine whether teachers are able to successfully encourage students to read more in ways that would actually improve fluency and overall reading ability. For the most part, these studies emphasize silent reading procedures, that is, students reading individually on their own with little or no specific feedback. Although the immediate impact of encouraging students to read would be expected first to increase the amount of reading engaged in, then to improve fluency in the ways discussed earlier, and finally to improve comprehension, that is not how these studies have been conducted. Studies of encouraging students to read rarely measure the actual increase in amount of reading due to the encouragement procedures, and they measure only the ultimate outcome (i.e., improvement in reading comprehension) rather than the intermediary enhancement to fluency that would be expected from the increased practice.

Encouraging Students to Read More: Methodology

Database

As with the search on repeated reading and guided oral reading, it is important to proceed in a systematic, replicable way and to describe these procedures thoroughly so that others can examine this work critically.

Consideration of Extant Literature Searches
This search started with the location of a published literature review on the impact of reading [Cunningham & Stanovich (1998). What reading does for the mind. American Educator, 22(1-2), 8-15.] This paper was examined carefully to identify appropriate terminology that could be used to conduct a thorough electronic search of the literature, and the reference list from that study was examined for additional, potentially relevant studies on this topic.

Identification of Appropriate Terminology
This search used electronic databases, which require appropriate search terms. In addition to conducting this literature review, the Panel examined various published reference sources to help identify terms for use in the search. The Panel used The Literacy Dictionary (Harris & Hodges, 1995); Handbooks of Reading Research I and II (Barr, Kamil, Mosenthal, et al., 1991; Pearson, Barr, Kamil, et al., 1984); The Encyclopedia of English Studies and Language Arts (Purves, 1994); and the Handbook of Research on Teaching the English Language Arts (Flood, Jensen, Lapp, et al., 1991). The sources were examined for articles on uninterrupted sustained silent reading, reading preferences and interests, Matthew effects, voluntary reading, and other relevant topics identified during this analysis and from the literature search.

These efforts led to the identification of terms generally related to the concept of increased reading as well as to specific instructional approaches used for that purpose. Table 2 provides a listing of the 30 search terms and names that were used in this synthesis.

Table 2

Terms used to search the electronic databases for studies that encouraged student reading.

- free reading
- voluntary reading
- SSR
- USSR
- SQUIRT
- DEAR
- reading volume
- summer reading
- reading amount
- book flood
- community literacy
- Accelerated Reader
- leisure time
- Magazine Recognition Test
- Input hypothesis
- recreational reading
- independent reading
- sustained silent reading
- uninterrupted sustained silent reading
- super quiet reading time
- Matthew effects
- volume of reading
- reading time
- amount of reading
- leisure reading
- self selection
- choice behavior
- Author Recognition Test
- free voluntary reading
- Stephen Krashen

Electronic Search Strategies
Because of the nature of the topic and the possibility that a single search could miss key information, the Panel examined two separate databases: ERIC and PsycINFO. The Panel searched PsycINFO using the terminology listed in Table 2. Each of these terms was linked by OR statements, meaning that if any article in that database focused on any of these topics it would be included in our target pool. The target pool that the Panel identified in this way included 18,990 articles. Then a separate focus pool was constructed using the terms: reading, reading ability, reading achievement, reading comprehension, reading development, reading disabilities, reading education, reading materials, reading, reading measures, reading readiness, reading skills, reading speed, remedial reading, and silent reading. These reading topics were linked with each other by OR, again with the idea of identifying all articles about any aspect of reading in the PsycINFO database. The focus pool included 34,448 articles. This focus pool was then combined with the target pool using AND as the link. This means that the Panel was discarding anything in the target pool that was not clearly linked with reading or reading education. The resulting combination resulted in the identification of 1,021 potential articles; once non-English language

articles were deleted, 909 articles remained. Because this was judged to be too many to search for, the Panel limited the search to 1991 (inclusive) and identified 478 potential articles in the intersection of the target and focus pools for those years.

Next the Panel completed a similar search of the ERIC system. The Panel used all the terms listed in Table 2 to develop a target pool. This resulted in the identification of 5,645 possible articles published since 1984. The Panel then developed a focus pool using the terms: basal reading, beginning reading, content area reading, corrective reading, critical reading, decoding, directed reading activity, early reading, functional reading, independent reading, individualized reading, informal reading inventories, reading, reading ability, reading achievement, reading assignments, reading attitudes, reading comprehension, reading difficulties, reading failure, reading habits, reading improvement, reading instruction, reading interests, reading material selection, reading materials, reading motivation, reading processes, reading programs, reading rate, reading research, reading skills, reading strategies, recreational reading, remedial reading, silent reading, story reading, supplementary reading materials, OR sustained silent reading. There were 38,799 potential articles in the focus pool that included 1984. These were then crossed with the target pool, and this led to the identification of 1,669 potential articles, which were then limited to journal articles written in the English language (655 articles), with 325 of these published since 1991.

Analysis

The NRP combined the two searches to eliminate duplication and found 603 unique articles on these topics as a result of the two searches. Each abstract was read and coded to determine whether to include it in this analysis. The criteria for inclusion were that:

1. The study had to be a research study that appeared to consider the effect of encouraging students to read more on reading achievement.
2. The study had to focus on English reading education, conducted with children (K-12).
3. The study itself had to have appeared in a refereed journal.
4. The study had to be have been carried out with English language reading.

If an article was clearly inappropriate in terms of these criteria, it was rejected without search. Rejected articles were designated as (1) nonrefereed, (2) nonresearch, (3) off topic/off sample, or (4) non-English language instruction. Although an abstract might have had several violations, only one needed to be noted for an article to be rejected. As a result of this screening, the Panel attempted to locate 92 articles for further consideration.

Location of Articles

Of the 92 articles on encouraging students to read more, the Panel was able to locate 82, or 89% of the total. Each of the 79 articles that was located was reviewed to determine its relevance to the topic and its adherence to the various selection criteria. Any study that appeared to meet the criteria was then coded for possible use in the final analysis. Only nine papers survived this review because most of these turned out to be correlational studies that just attempted to test whether better readers read more, something that the Panel accepts as already proven.

Additional Identification of Articles

The Panel's search procedures neglected older studies of these instructional procedures. Only studies published since 1991 had been included in the selection procedures up to this point. To expand on this set of studies in an effective and efficient manner, the Panel analyzed the reference lists of all studies that were located through the previously described procedures. Even studies that were determined to be in violation of the final selection criteria were analyzed in this way. This led to the consideration of 46 additional papers, and of these, the Panel was able to locate 42 or 91%. For the most part, these second-generation papers had been published before 1990. Of the 42 papers, 10 appeared to meet all of the selection criteria. These 10 studies were added to the 9 previously identified, and these were designated for further examination and coding.

On closer examination, the Panel discovered that five of these studies were actually correlational studies and not experimental studies. This left only 14 studies with potential for answering this question.

Consistency With NRP Methods

The methods of the NRP were followed in the conduct of the literature searches and the examination of the articles obtained. However, in the case of these 14 studies, the Panel quickly realized that there were very few papers. Furthermore, the Panel evaluated a variety of procedures and found that many of the papers suffered from especially weak research design. Several of these 14 studies, although they met the selection criteria, could not be analyzed because of serious methodological or reporting flaws that undermined their results. Because of these concerns, the Panel did not think it appropriate to carry out a meta-analysis of the data. The Panel's concern was that the meta-analysis would be potentially misleading given the very limited data set that would be used for the analysis. Thus, this set of studies prohibited the of the NRP criteria for multiple studies.

Encouraging Students to Read More: Results and Discussion

Description of the Studies

Given that only 14 studies fit the selection criteria, it seems reasonable to summarize each one. The studies are listed in Appendix E. Most of the 14 studies examined the impact of sustained silent reading (SSR), but some other approaches were also studied. SSR goes under a variety of labels including USSR (uninterrupted sustained silent reading), DEAR (drop everything and read), and SQUIRT (super quiet reading time). In most cases, these procedures require the provision of approximately 20 minutes per day in which students are allowed to read material silently on their own with no monitoring. In most cases, the students select their own material, and there is no discussion or written assignment tied to this reading. Teachers and other adults in the school setting are to read during this time as well. Such programs are described in nearly all teacher preparation textbooks and have become widely popular in American classrooms in both elementary and secondary schools.

Sustained Silent Reading (SSR)

One study of SSR (Evans & Towner, 1975) compared the effect of SSR on reading achievement with that of having students complete various reading skills exercises with commercial materials (i.e., worksheets). Reading gains were identical for both groups of 2nd graders at the end of 10 weeks.

In a similar, though larger study, Reutzel and Hollingsworth (1991) compared skills practice and SSR with 61 4th graders and 53 6th graders. These procedures were used for 1 month, and there were, again, no reading differences for the two approaches. As with the previous study, the skills work was assembled by the researchers specifically to serve as a control activity, and was not part of the regular instructional program that these students received from their teachers.

Collins (1980) conducted an analysis of the impact of SSR on the reading achievement of 220 students from ten classrooms in grades 2 through 6. Students were randomly assigned to the experimental and control groups. This daily program was evaluated after 15 weeks (different grade levels allotted different amounts of time to SSR—2nd graders had 10 to 30 minutes per day; 3rd graders received 15 minutes daily; 4th graders, 30 minutes; and 5th and 6th graders, 15 to 25 minutes each day). The control group worked on spelling during these time periods. The SSR procedures led to no significant differences in vocabulary or comprehension as measured by various standardized tests, although the SSR groups appeared to move slightly faster through their basal readers during this period.

Langford and Allen (1983) examined the impact of SSR on the reading attitudes and achievement of 11 5th and 6th grade classes. These classes were randomly assigned to SSR or control conditions, resulting in 131 students in the SSR group (60 5th graders and 71 6th graders) and 119 students in the control group. Students in the control group learned about health and grooming while the SSR activities took place with the experimental subjects. The study failed to report the length of the instructional period or the duration of the intervention. Although there was significantly better improvement in word reading for the SSR group, these differences appear to be small in terms of educational importance. In any event, it is difficult to evaluate the value of these gains without more information about the length of the program. There were no differences in reading attitude that resulted from the intervention.

In still another evaluation of SSR, this one conducted in a junior high school, Cline and Kretke (1980) examined the effectiveness of the procedure over a 3-year period. This study compared the reading achievement of 111 students who had been enrolled for 3 years at a junior high school that was using SSR with that of control group students drawn from two other schools that did not have this program. This study found no differences between the two groups. However, it was poorly designed, and it would be impossible to be certain whether there were gains. The study apparently compared gains between different achievement tests used at different grade levels (something that is not statistically sound), and it failed to provide any information about the length of the SSR time or how this time was used at the control school.

Davis (1988) considered the effect of SSR on reading comprehension with 8th graders. Fifty-six students were randomly assigned to one of two English classes. These classes met daily for 50 minutes. Approximately half the time was devoted to either SSR or, alternatively, to directed reading activities with the teacher. This effort continued for an entire school year. Although the researcher intended to analyze these data for high-, medium-, and low-ability students separately, attrition in the low-ability groups rendered this impossible. Two comparisons were made for the high- and medium-ability groups, and it was found that the medium-ability students made much greater gains with SSR than with directed reading (n = 19), but there were no significant differences among the two high-ability groups (15 students in these two groups). The gains attributed to SSR for the medium-ability group were substantial and educationally meaningful (about 1 year of difference on a standardized test). Unfortunately, the study is somewhat sketchy in terms of the statistical analysis: it provided no means or standard deviations and told little about the analysis of covariance that was used (i.e., How big were the initial differences across the groups? Was heterogeneity tested?).

In one of the best-designed studies on SSR, Holt and O'Tuel (1989) randomly assigned teachers and 211 7th and 8th grade students to an SSR condition and a regular reading instruction condition. Students in the SSR condition read self-selected materials for 20 minutes per day for 3 days each week, and they carried out sustained silent writing for two additional 20-minute periods each week. During the time these activities were carried out, the control group subjects worked on their regular reading instruction. At the end of 10 weeks, the students in the SSR groups had evidenced greater growth in vocabulary knowledge than was true for the control subjects. Reading comprehension did not improve for either group, however.

Burley (1980) randomly assigned 85 high school students enrolled in an Upward Bound summer program at a local college to one of four groups: SSR, programmed textbooks, programmed cassette tapes, and programmed skill development kits. The students in all groups received 75 minutes of reading instruction per day for 30 days, but part of this time was devoted to the SSR or other practice activities. In all, students practiced reading for about 14 hours in addition to the summer reading instruction during this 6-week period. This study found a small, positive, statistically significant difference favoring SSR over the other procedures on reading comprehension but no differences on a vocabulary measure.

Summers and McClelland (1982) examined the effect of a 5-month program of SSR with 65 intact treatment and control classes from nine elementary schools. They found no significant differences in covariance-adjusted mean scores from standardized and informal reading achievement and attitude measures and no significant interaction effects for reading achievement, attitude, grade level, and sex. This study included approximately 1,400 children. This study was unique not only in terms of its extensive sample, but also in that it carefully monitored the delivery of the treatments.

In yet another study of SSR (Manning & Manning, 1984), three variations of SSR were tested with 4th graders. These variations were compared across an entire school year with a poorly described control group. Students (n = 415) from 24 classrooms were assigned to the four groups (intact classes were randomly assigned). The treatment lasted for an entire school year. This study found that two of the SSR variations led to higher reading achievement and that one did not. The pure SSR variation (i.e., the one that matched the recommended procedures), in which students read for an extra 35 minutes per day, led to no greater reading growth than was evident for the control group. However, when SSR was coupled with teacher conferences or peer discussion, then slight improvement in reading was evident for the SSR groups. This suggests that reading alone might provide no clear benefit but that additional reading in combination with other activities could be effective.

Not all the studies in this category focused on SSR, however. Morrow and Weinstein (1986), for instance, worked with six 2nd-grade reading classes to determine the efficacy of being involved in either a home- or school-based voluntary reading program in terms of amount of reading and reading achievement. This program, which provided students with enriched library materials and extended reading time, lasted for 9 weeks. Students did more school reading as a result of being in this program, and they continued to do so when the program ended, but achievement levels in reading were unrelated to program participation, and the program did not alter reading attitudes or the amount of home reading.

Accelerated Reader (AR)

AR is a commercial program designed to increase the amount of reading that students do with appropriate materials. Peak and Dewalt (1994) compared reading gains for two schools, one that used this program and one that did not. To make this comparison, they randomly selected 50 9th graders from each school. To be selected, a student had to have attended these schools since grade 3. Because standardized reading test scores (California Achievement Test) were available for each school at 3rd, 6th, and 8th grades, comparisons were made between these two groups at each point. They found a slight reading advantage in 3rd grade scores for the school that did not use AR and a slight advantage for the AR group at the end of the year. Students in the AR group had taken part in 5 to 6 hours per week of in-class reading during the 5 years of this study, but there is no information on what the other students were doing during this time. More problematic is the calculation of gain scores across forms of a standardized test. The scores of each of these normative grade level tests are independent scales, and it is not valid to subtract these test scores from each other. Given this serious problem and the limited data reporting that was evident, it is unclear whether any real difference in achievement can be attributed to this program on the basis of this study.

In another study of the Accelerated Reader (Vollands, Topping, & Evans, 1999), two small experiments were carried out. In one experiment, there was a small advantage due to participation in the program; in the other, there was not. Neither study had well-matched samples of students, and in the study that demonstrated an advantage, students also used a form of assisted reading similar to those examined earlier in this paper.

Carver and Liebert (1995) provided one of the clearest tests of the effect of reading by studying students during the summer. This study did not have a control group but simply examined the reading scores at the beginning of the program and 6 weeks later after the students had completed approximately 60 hours of self-selected reading. These students, in 3rd through 5th grades, made no gains in reading achievement at all, even though the books were at an appropriate level.

Encouraging Students to Read More: Implications for Reading Instruction

None of these studies attempted to measure the effect of increased reading on fluency. Instead, most of these studies considered the impact of encouraging more reading on overall reading achievement as measured by standardized and informal tests. It would be difficult to interpret this collection of studies as representing clear evidence that encouraging students to read more actually improves reading achievement. Only three studies (Burley, 1980; Davis, 1988; Langford & Allen, 1983) reported any clear reading gains from encouraging students to read, and in the third of these studies the gains were so small as to be of questionable educational value. Most of the studies, including the best designed and largest ones (Collins, 1980; Holt & O'Tuel, 1989; Summers & McClelland, 1982), reported no appreciable benefit to reading from such procedures (Holt & O'Tuel found improvement in vocabulary scores, but these did not translate into better reading comprehension). The most direct test of the effect of reading on learning was provided by Carver and Liebert

231

(1995), and they found no clear benefit resulting from 60 hours of additional reading. Perhaps 60 hours of reading is insufficient for improving achievement in a measurable way.

Only two of the studies compared SSR with nonreading instruction (Collins, 1980; Langford & Allen, 1983). One of these found no benefit, and the other found a very small benefit from SSR. More of the studies compared additional reading time with reading instruction itself. Often these studies interpreted the lack of difference between SSR and the control condition as meaning that SSR was as good as some, usually unspecified, form of reading instruction. Comparing SSR with instructional routines that have no evidence of success—or whose success has been found to be unrelated to achievement gains (Leinhardt, Zigmond, & Cooley, 1981)—is meaningless. Although several reviews of the literature have concluded that procedures like SSR work simply because reading achievement does not decline once they are instituted, that is not a sound basis on which to recommend such procedures as effective. SSR may or may not work, but it is unreasonable to conclude that it does on the basis of such flawed reasoning. For the most part, these studies found no gains in reading due to encouraging students to read more. It is unclear whether this was the result of deficiencies in the instructional procedures themselves or to the weaknesses and limitations evident in the study designs.

It is impossible to sustain a negative conclusion with research. That is, the NRP cannot ultimately prove that a procedure or approach does not work under any conditions. No matter how many studies show a lack of effect due to an instructional routine, it is always possible that under some yet-unstudied condition the procedure could be made to work. Given the paucity of studies on increasing the amount of student reading—and the uneven quality of much of this work—there is a need to be especially cautious. Few of the studies reviewed here provided much monitoring of the amount of reading that students actually did in the programs, and only one kept track of the control student reading; therefore, in most cases, it is unclear whether the interventions actually led to more reading or just displaced other reading that students might have done otherwise. Nevertheless, given the evidence that exists, the Panel cannot conclude that schools should adopt programs to encourage more reading if the intended goal is to improve reading achievement. It is not that studies have proven that this cannot work, only that it is yet unproven.

There are few beliefs more widely held than that teachers should encourage students to engage in voluntary reading and that if they did this successfully, better reading achievement would result. Unfortunately, research has not clearly demonstrated this relationship. In fact, the handful of experimental studies in which this idea has been tried raise serious questions about the efficacy of some of these procedures.

Encouraging Students to Read More: Directions for Further Research

There is a need for rigorous evaluations of the effectiveness of encouraging wide reading on reading achievement, particularly with popular programs such as SSR, DEAR, and AR. These studies need to monitor the amounts of reading—in and out of school—by both the experimental and control group students. To really understand the implications of such reading, it is important to compare these routines against procedures in which students actually read less. Without such information, one might only be comparing the effects of different forms of reading practice rather than comparing differences in amount of reading practice. Finally, none of these studies could even demonstrate that they clearly increased the amount of student reading because none of them measured an adequate baseline of current or previous reading engagement. That, too, should be addressed in future studies.

That encouraging more reading does as well as certain instructional activities in stimulating learning does not speak well of those instructional activities. Voluntary reading within the school day should be compared against nonreading activities or activities in which the amount of reading can be closely measured. (In fact, the field should consider adopting a new research convention for methodological studies with students in the 2nd grade or higher. The amount of gain attributable to reading alone should be the baseline comparison against which the efficacy of instructional procedures is tested. If an instructional method does better than reading alone, it would be safe to conclude that method works.) Studies should consider the effect of increasing student reading on both fluency and overall reading achievement. However, until such evidence is forthcoming, the National Reading Panel cannot indicate that research has proven that such procedures actually work.

Overall Conclusions

Fluency is an essential part of reading, and the NRP has reviewed its theoretical and practical implications for reading development. In addition, the Panel has conducted two research syntheses, one on guided oral reading procedures such as repeated reading and the other on the effect of procedures that encourage students to read more. These two procedures have been widely recommended as appropriate and valuable avenues for increasing fluency and overall reading achievement.

The NRP found a better, and more extensive, body of research on guided oral reading procedures. Generally, the Panel found that these procedures tended to improve word recognition, fluency (speed and accuracy of oral reading), and comprehension with most groups. Although there has been some speculation that fluency development is complete for most students by grade 3 or 4, the Panel's analysis found that these procedures continue to be useful far beyond that—at least for some readers. Repeated reading and other guided oral reading procedures have clearly been shown to improve fluency and overall reading achievement. There is clear and substantial research evidence that shows that such procedures work under a wide variety of conditions and with minimal special training or materials. Even with this evidence, there is a need for more research on this topic, including longitudinal studies that examine the impact of these procedures on different levels of students over longer periods. It would also be worthwhile to determine the amount of such instruction that would be needed with most students and the types of materials that lead to the biggest gains when these procedures are used.

The results of the analysis of programs that encourage students to read more were much less encouraging. Despite widespread acceptance of the idea that schools can successfully encourage students to read more and that these increases in reading practice will be translated into better fluency and higher reading achievement, there is not adequate evidence to sustain this claim. Few studies have attempted to increase the amount of student reading. Those that have investigated such issues have tended to find no gains in reading as a result of the programs. This does not mean that procedures that encourage students to read more could not be made to work—future studies should explore this possibility—but at this time, it would be unreasonable to conclude that research shows that encouraging reading has a beneficial effect on reading achievement.

References

Advantage Learning Systems. (1986). Accelerated Reader. Wisconsin Rapids, WI: Advantage Learning Systems.

Ackerman, P. L. (1987). Individual differences in skill reading: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Allington, R. (1984). Content coverage and contextual reading in reading groups. Journal of Reading Behavior, 16, 85-96.

Allington, R. L. (1983). Fluency: The neglected reading goal in reading instruction. The Reading Teacher, 36, 556-561.

Allington, R. L. (1977). If they don't read much, how they ever gonna get good? Journal of Reading, 21, 57-61.

Anderson, R. C., Wilkinson, I. A. G., & Mason, J. M. (1991). A microanalysis of the small-group, guided reading lesson: Effects of an emphasis on global story meaning. Reading Research Quarterly, 26, 417-441.

Barr, R., Kamil, M. L., Mosenthal, P., & Pearson, P. D. (1991). Handbook of reading research (Vol. 2). New York: Longman.

Biemiller, A. (1977-78). Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. Reading Research Quarterly, 13, 223-253.

Bryan, W. L., & Harter, N. (1899). Studies of the telegraphic language. The acquisition of a hierarchy of habits. Psychological Review, 6, 345-375.

Buswell, G. T. (1922). Fundamental reading habits: A study of their development. Supplementary Educational Monographs, No. 21. Chicago: University of Chicago Press.

Buswell, G. T. (1937). How adults read. (Supplementary Educational Monographs). Chicago: University of Chicago Press.

Calfee, R. C., & Piaotkowski, D. C. (1981). The reading diary: Acquisition of decoding. Reading Research Quarterly, 16, 346-373.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), Eye movements in reading: Perceptual and language processes (pp. 275-307). New York: Academic Press.

Cattell, J. M. (1885). Uber die zeit der erkennung und bennenung von schriftzeichen. Bildern und Farben Studien, 2, 635-650.

Clay, M. M. (1972). The early detection of reading difficulties. Auckland, NZ: Heinemann.

Cooper, H. (1998). Synthesizing research (3rd ed.). Thousand Oaks, CA: Sage.

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. American Educator, 22(1-2), 8-15.

Donahue, P. L., Voelkl, K. E., Campbell, J. R., & Mazzeo, J. (1999). NAEP 1998 reading report card for the nation and states. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. Reading Research Quarterly, 22, 389-406.

Dowhower, (1994). Repeated reading revisited: Research into practice. Reading and Writing Quarterly, 10, 343-358.

Everatt, J., Bradshaw, M. F., & Hibbard, P. B. (1998). Individual differences in reading and eye movement control. In G. Underwood (Ed.), Eye guidance in reading and scene perception (pp. 223-242). Oxford, England: Elsevier.

Everatt, J., & Underwood, G. (1994). Individual differences in reading subprocesses: Relationships between reading ability, lexical access, and eye movement control. Language and Speech, 37, 283-297.

Faulkner, H. J., & Levy, B. A. (1999). Fluent and nonfluent forms of transfer in reading: Words and their message. Psychonomic Bulletin and Review, 6, 111-116.

Fleischer, L. S., Jenkins, J., & Pany, D. (1979). Effects on poor readers' comprehension of training in rapid decoding. Reading Research Quarterly, 14, 30-48.

Flood, J., Jensen, J. M., Lapp, D., & Squire, J. R. (1991). Handbook of research on teaching the English language arts. New York: Macmillan.

Frazier, L., & Rayner, K. (1982). Making & correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178-210.

Fries, C. C. (1962). Linguistics and reading. New York: Holt.

Goodman, Y. M., & Burke, C. L. (1972). Reading miscue inventory: Procedure for diagnosis and correction. New York: Macmillan.

Gough, P. B. (1972). One second of reading. In J. F. Kavanagh, & I. G. Mattingly (Eds.), Language by ear and by eye. (pp. 331-358). Cambridge, MA: MIT Press.

Greene, F. P. (1979). Radio reading. In C. Pennock (Ed.), Reading comprehension at four linguistic levels (pp. 104-107). Newark, DE: International Reading Association.

Harris, T. L., & Hodges, R. E. (1995). The literacy dictionary. Newark, DE: International Reading Association.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. Teaching Exceptional Children, 24(3), 41-44.

Heckelman, R. G. (1969). A neurological-impress method of remedial-reading instruction. Academic Therapy, 4, 277-282.

Herman, P. A. (1985). The effect of repeated readings on reading rate, speech pauses, and word recognition accuracy. Reading Research Quarterly, 20, 553-565.

Homan, S. P., Klesius, J. P., & Hite, C. (1993). Effects of repeated readings and nonrepetitive strategies on students' fluency and comprehension. Journal of Educational Research, 87, 94-99.

Huey, E. B. (1905). The psychology and pedagogy of reading. Cambridge, MA: MIT Press.

Hunt, L. C., Jr. (1970). The effect of self-selection, interest, and motivation on independent, instructional, and frustrational levels. The Reading Teacher, 24, 146-151, 158.

Ikeda, M., & Saida, S. (1978). Span of recognition in reading. Vision Research, 18, 83-88.

James, W. (1890). The principles of psychology. New York: Holt.

Johnson, M. S., Kress, R. A., & Pikulski, J. J. (1987). Informal reading inventories (2nd ed.). Newark, IL: International Reading Association.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. Journal of Educational Psychology, 80, 437-447.

Kennedy, A. (1983). On looking into space. In K. Rayner (Ed.), Eye movements in reading: Perceptual and language processes (pp. 237-251). New York: Academic Press.

Kennedy, A., & Murray, W. S. (1987a). The components of reading time: Eye movement patterns of good and poor readers. In J. K. O'Reagan & A. Levy-Schoen (Eds.), Eye movements: From physiology to cognition (pp. 509-520). Amsterdam: North Holland.

Kennedy, A., & Murray, W. S. (1987b). Spatial coordinates and reading: Comments on Monk. Quarterly Journal of Experimental Psychology, 39A, 649-656.

Krashen, S. D. (1993). The power of reading: Insights from the research. Englewood, CO: Libraries Unlimited.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Leinhardt, G., Zigmond, N., & Cooley, W. W. (1981). Reading instruction and its effects. American Educational Research Journal, 18, 343-361.

Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. Journal of Experimental Child Psychology, 56, 303-327.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. Reading and Writing Quarterly, 13, 123-146.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. Perception and Psychophysics, 17, 578-586.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? Perception and Psychophysics, 25, 221-224.

McConkie, G. W., & Zola, D. (1987). Visual attention during eye fixations while reading. In M. Coltheart (Ed.), Attention and performance (Vol. 12, pp. 385-401). London: Erlbaum.

Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. Quarterly Journal of Experimental Psychology, 40A, 693-718.

Nagy, W., & Anderson, R. C. (1984). How many words are there in printed school English? Reading Research Quarterly, 19, 304-330.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. Journal of Experimental Psychology: General, 106, 226-254.

Neill, K. (1979). Turn kids on with repeated reading. Teaching Exceptional Children, 12, 63-64.

Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual differences and developmental differences in reading disability. In G. MacKinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (pp. 1-64). New York: Academic Press.

Opitz, M. F., & Rasinski, T. V. (1998). Good-bye round robin. Portsmouth, NH: Heinemann.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. Journal of Reading Behavior, 17, 129-142.

Pearson, P. D., Barr, R., Kamil, M. L., & Mosenthal, P. (1984). Handbook of reading research. New York: Longman.

Pflaum, S. W., & Pascarella, E. T. (1980). Interactive effects of prior reading achievement and training in context on the reading of learning-disabled children. Reading Research Quarterly, 16, 138-158.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). Listening to children read aloud. Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education.

Posner, M. I., & Snyder, R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition: The Loyola Symposium (pp. 55-85). Hillsdale, NJ: Erlbaum.

Purves, A. C. (1994). Encyclopedia of English studies and language arts. New York: Scholastic.

Radach, R., & Kempe, V. (1993). An individual analysis of initial fixation positions in reading. In G. d'Ydewalle & J. Van Rensbergen (Eds.), Perception and cognition: Advances in eye movement research (pp. 213-226). Amsterdam: North Holland.

Rasinski, T. V. (1990). Effects of repeated reading and listening-while-reading on reading fluency. Journal of Educational Research, 83, 147-150.

Rayner, K. (1978b). Eye movements in reading and information processing. Psychological Bulletin, 85, 618-660.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. Journal of Experimental Child Psychology, 41, 211-236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372-422.

Rayner, K., & Duffy, S. A. (1988). On-line comprehension processes and eye movements in reading. In M. Daneman, G. E. MacKinnon, & T. G. Waller (Eds.), Reading research: Advances in theory and practice (pp. 13-66). San Diego, CA: Academic Press.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. Cognitive Psychology, 12, 206-226.

Rayner, K., & Pollatsek, A. (1994). The psychology of reading. Mahwah, NJ: Erlbaum.

Samuels, S. J. (1979). The method of repeated readings. The Reading Teacher, 32, 403-408.

Samuels, S. J., LaBerge, D., & Bremer, C. (1978). Units of word recognition: Evidence for developmental changes. Journal of Verbal Learning and Verbal Behavior, 17, 715-720.

Samuels, S. J., Miller, N., & Eisenberg, P. (1979). Practice effects on the unit of word recognition. Journal of Educational Psychology, 71, 514-520.

Schreiber, P. A. (1980). On the acquisition of reading fluency. Journal of Reading Behavior, 12, 177-186.

Schreiber, P. A. (1987). Prosody and structure in children's syntactic processing. In R. Horowitz & S. J. Samuels (Eds.), Comprehending oral and written language. New York: Academic Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic information processing: Perceptual learning, automatic attending, and general theory. Psychological Review, 96, 84, 127-190.

Shanahan, S., Wojciehowski, J., & Rubik, G. (1998). A celebration of reading: How our school read for one million minutes. The Reading Teacher, 52, 93-96.

Shaywitz, B., Holford, T.R., Holahan, J.M., Fletcher, J.M., Steubing, K.K., Francis, D.J., & Shaywitz, S.E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study of reading. Reading Research Quarterly, 30, 894-906.

Sindelar, P. T., Monda, L. E., & O'Shea, L. J. (1990). Effects of repeated readings on instructional- and mastery-level readers. Journal of Educational Research, 83, 220-226.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 8(11), 11-16.

Stanovich, K. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. Developmental Review, 10, 72-100.

Stanovich, K. (1986). Mathew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading Research Quarterly, 22, 360-406.

Strecker, S., Roser, N., & Martinez, M. (1998). Toward understanding oral reading fluency. In T. Shanahan & F. Rodriguez-Brown (Eds.) Forty seventh Yearbook of the National Reading Conference, (pp. 295-310). Chicago, IL: The National Reading Conference.

Taylor, B. M., Pearson, P. D., Clark, K. F., & Walpole, S. (1999). Beating the odds in teaching all children to read. (CIERA Rep. No. 2-006). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement, University of Michigan.

Taylor, S., Frackenpohl, H., & Pettee, J. (1960). Grade level norms for the components of the fundamental reading skill. (Bulletin No. 3). Huntington, NY: Educational Developmental Laboratories.

Terry, P., Samuels, S. J., & LaBerge, D. (1976). The effects of letter degradation and letter spacing on word recognition. Journal of Verbal Learning and Verbal Behavior, 15, 577-585.

Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation. Reading and Writing Quarterly, 13, 165-184.

Topping, K. (1987). Paired reading: A powerful technique for parent use. The Reading Teacher, 40, 608-614.

Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationship between reading skill, reading speed and visual inspection. Language and Speech, 33, 69-81.

Underwood, N. R., & McConkie, G. W. (1985). Perceptual span for letter distinctions during reading. Reading Research Quarterly, 20, 153-162.

Van Bon, W. H. J., Boksebeld, L. M., Font Freide, T. A. M., & van den Hurk, A. J. M. (1991). A comparison of three methods of reading-while-listening. Journal of Learning Disabilities, 24, 471-476.

VanWagenen, M. A., Williams, R. L., & McLaughlin, T. F. (1994). Use of assisted reading to improve reading rate, word accuracy, and comprehension with ESL Spanish-speaking students. Perceptual and Motor Skills, 79, 227-230.

Wagner, R., Torgesen, J. & Rashotte, C. (1999). Comprehensive test of phonological processes. Austin, TX: Pro-Ed.

Wiederholt, J. L. & Bryant, B. R. (1992). Gray Oral Reading Tests. (3rd ed.). Austin, TX: Pro-Ed.

Winter, S. (1986). Peers as paired reading tutors. British Journal of Special Education, 13, 103-106.

Winter, S. (1988). Paired reading: A study of process and outcome. Educational Psychology, 8, 135-151.

FLUENCY

Appendices

Appendix A

Studies of Repeated Reading and Guided Oral Reading That Tested Immediate Impact of the Procedures on Reading Performance (no transfer)

Faulkner, H. J., & Levy, B. A. (1999). Fluent and nonfluent forms of transfer in reading: Words and their message. Psychonomic Bulletin and Review, 6, 111-116.

Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. Journal of Experimental Child Psychology, 56, 303-327.

Neill, K. (1979). Turn kids on with repeated reading. Teaching Exceptional Children, 12, 63-64.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. Journal of Reading Behavior, 17, 129-142.

Pany, D., & McCoy, K. M. (1988). Effects of corrective feedback on word accuracy and reading comprehension of readers with learning disabilities. Journal of Learning Disabilities, 21, 546-550.

Rasinski, T. V. (1990). Effects of repeated reading and listening-while-reading on reading fluency. Journal of Educational Research, 83, 147-150.

Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening, and independent reading with computer-based speech feedback. Reading Research Quarterly, 23, 219-235.

Rose, T. L., & Beattie, J. R. (1986). Relative effects of teacher-directed and taped previewing on oral reading. Learning Disability Quarterly, 9, 193-199.

Sindelar, P. T., Monda, L. E., & O'Shea, L. J. (1990). Effects of repeated readings on instructional- and mastery-level readers. Journal of Educational Research, 83, 220-226.

Smith, D. D. (1979). The improvement of children's oral reading through the use of teacher modeling. Journal of Learning Disabilities, 12 (3), 39-42.

Stoddard, K., Valcante, G., Sindelar, P., O'Shea, L., & Algozzine, B. (1993). Increasing reading rate and comprehension: The effects of repeated readings, sentence segmentation, and intonation training. Reading Research and Instruction, 32, 53-65.

Taylor, N. E., Wade, M. R., & Yekovich, F. R. (1985). The effects of text manipulation and multiple reading strategies on the reading performance of good and poor readers. Reading Research Quarterly, 20, 566-574.

Turpie, J. J., & Paratore, J. R. (1995). Using repeated reading to promote success in a heterogeneously grouped first grade. In K. A. Hinchman, D.J. Leu, & C.K. Kinzer (Eds.) Perspectives on literacy research and practice: Forty-fourth Yearbook of the National Reading Conference (pp. 255-263). Chicago: The National Reading Conference.

VanWagenen, M. A., Williams, R. L., & McLaughlin, T. F. (1994). Use of assisted reading to improve reading rate, word accuracy, and comprehension with ESL Spanish-speaking students. Perceptual and Motor Skills, 79, 227-230.

239

Appendix B

Articles Included in Meta-Analysis on Guided Oral Reading Procedures

Conte, R., & Humphreys, R. (1989). Repeated readings using audiotaped material enhances oral reading in children with reading difficulties. Journal of Communications Disorders, 22, 65-79.

Eldredge, J. L. (1990). Increasing the performance of poor readers in the third grade with a group-assisted strategy. Journal of Educational Research, 84, 69-77.

Eldredge, J. L., Reutzel, D. R., & Hollingsworth, P. M. (1996). Comparing the effectiveness of two oral reading practices: Round-robin reading and the shared book experience. Journal of Literacy Research, 28, 201-225.

Hollingsworth, P. M. (1978). An experimental approach to the impress method of teaching reading. Reading Teacher, 31, 624-627.

Hollingsworth, P. M. (1970). An experiment with the impress method of teaching reading. Reading Teacher, 24, 112-114, 187.

Labbo, L. D., & Teale, W. (1990). Cross-age reading: A strategy for helping poor readers. Reading Teacher, 43, 362-369.

Lorenz, L., & Vockell, E. (1979). Using the neurological impress method with learning disabled readers. Journal of Learning Disabilities, 12, 67-69.

Mathes, P. G., & Fuchs, L. S. (1993). Peer mediated reading instruction in special education resource rooms. Learning Disabilities Research and Practice, 8, 233-243.

Miller, A., Robson, D., & Bushell, R. (1986). Parental participation in paired reading: A controlled study. Educational Psychology, 6, 277-284.

Rasinski, T., Padak, N., Linek, W., & Sturtevant, E. (1994). Effects of fluency development on urban second-grade readers. Journal of Educational Research, 87, 158-165.

Reutzel, D. R., & Hollingsworth, P. M. (1993). Effects of fluency training on second graders' reading comprehension. Journal of Educational Research, 86, 325-331.

Shany, M. T., & Biemiller, A. (1995). Assisted reading practice: Effects on performance for poor readers in grade 3 and 4. Reading Research Quarterly, 30, 382-395.

Simmons, D., Fuchs, D., Fuchs, L. S., Hodge, J. P., & Mathes, P. G. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. Learning Disabilities Research and Practice, 9, 203-212.

Simmons, D.C., Fuchs, L. S., Fuchs, D., Mathes, P., & Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. Elementary School Journal, 95, 387-408.

Thomas, A., & Clapp, T. (1989). A comparison of computer-assisted component reading skills training and repeated reading for adolescent poor readers. Canadian Journal of Special Education, 5, 135-144.

Young, A. R., Bowers, P. G., & MacKinnon, G. E. (1996). Effects of prosodic modeling and repeated reading on poor readers' fluency and comprehension. Applied Psycholinguistics, 17, 59-84.

Appendix C

Studies of Guided Oral Reading That Used Single Subject Designs

Blum, I. H., Koskinen, P. S., Tennant, N., Parker, E. M., Straub, M., & Curry, C. (1995). Using audiotaped books to extend classroom literacy instruction into the homes of second-language learners. Journal of Reading Behavior, 27, 535-563.

Gilbert, L. M., Williams, R. L., & McLaughlin, T. F. (1996). Use of assisted reading to increase correct reading rates and decrease error rates of students with learning disabilities. Journal of Applied Behavior Analysis, 29, 255-257.

Herman, P. A. (1985). The effect of repeated readings on reading rate, speech pauses, and word recognition accuracy. Reading Research Quarterly, 20, 553-565.

Kamps, D. M., Barbetta, P. M., Leonard, B. R., & Delquadri, J. (1994). Classwide peer tutoring: An integration strategy to improve reading skills and promote peer interactions among students with autism and general education peers. Journal of Applied Behavior Analysis, 27, 49-61.

Langford, K., Slade, K., & Barnett, A. (1974). An examination of impress techniques in remedial reading. Academic Therapy, 9, 309-319.

Law, M., & Kratochwill, T. R. (1993). Paired reading: An evaluation of a parent tutorial program. School Psychology International, 14, 119-147.

Mefferd, P. E., & Pettegrew, B. S. (1997). Fostering literacy acquisition of students with developmental disabilities: Assisted reading with predictable trade books. Reading Research and Instruction, 36, 177-190.

Morgan, R. T. (1976). "Paired reading" tuition: A preliminary report on a technique for cases of reading deficit. Child: Care, Health and Development, 2, 13-28.

Morgan, R., & Lyon, E. (1979). "Paired reading"–A preliminary report on a technique for parental tuition of reading-retarded children. Journal of Child Psychiatry, 20, 151-160.

Rose, T. L. (1984). The effects of two prepractice procedures on oral reading. Journal of Learning Disabilities, 17, 544-548.

Tingstrom, D. H., Edwards, R. P., & Olmi, D. J. (1995). Listening previewing in reading to read: Relative effects on oral reading fluency. Psychology in the Schools, 32, 318-327.

Weinstein, G., & Cooke, N. L. (1992). The effects of two repeated reading interventions on generalization of fluency. Learning Disability Quarterly, 15, 21-28.

241

Appendix D

Studies That Compared Methods of Guided Oral Reading

Carver, R. P., & Hoffman, J. V. (1981). The effect of practice through repeated reading on gain in reading ability using a computer-based instructional system. Reading Research Quarterly, 3, 374-390.

Dixon-Krauss, L. A. (1995). Partner reading and writing: Peer social dialogue and the zone of proximal development. Journal of Reading Behavior, 27, 45-63.

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. Reading Research Quarterly, 22, 389-406.

Homan, S. P., Klesius, J. P., & Hite, C. (1993). Effects of repeated readings and nonrepetitive strategies on students' fluency and comprehension. Journal of Educational Research, 87, 94-99.

Lindsay, G., Evans, A., & Jones, B. (1985). Paired reading versus relaxed reading: A comparison. British Journal of Educational Pscyhology, 55, 304-309.

Rashotte, C., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. Reading Research Quarterly, 20, 180-188.

Van Bon, W. H. J., Boksebeld, L. M., Font Freide, T. A. M., & van den Hurk, A. J. M. (1991). A comparison of three methods of reading-while-listening. Journal of Learning Disabilities, 24, 471-476.

Winter, S. (1986). Peers as paired reading tutors. British Journal of Special Education, 13, 103-106.

Winter, S. (1988). Paired reading: A study of process and outcome. Educational Psychology, 8, 135-151.

242

Appendix E

Studies of the Effects of Encouraging Students to Read

Burley, J. E. (1980). Short-term, high intensity reading practice methods for Upward Bound Students: An appraisal. Negro Educational Review, 31, 156-161.

Carver, R. P., & Liebert, R. E., (1995). The effect of reading library books in different levels of difficulty on gain in reading ability. Reading Research Quarterly, 30, 26-48.

Cline, R. K. J., & Kretke, G. L. (1980). An evaluation of long-term SSR in the junior high school. Journal of Reading, 23, 503-506.

Collins, C. (1980). Sustained silent reading periods: Effects on teachers' behaviors and students' achievement. Elementary School Journal, 81, 108-114.

Davis, Z. T. (1988). A comparison of the effectiveness of sustained silent reading and directed reading activity on students' reading achievement. The High School Journal, 72(1), 46-48.

Evans, H. M., & Towner, J. C. (1975). Sustained silent reading: Does it increase skills? Reading Teacher, 29, 155-156.

Holt, S. B., & O'Tuel, F. S. (1989). The effect of sustained silent reading and writing on achievement and attitudes of seventh and eighth grade students reading two years below grade level. Reading Improvement, 26, 290-297.

Langford, J. C., & Allen, E. G. (1983). The effects of U.S.S.R. on students' attitudes and achievement. Reading Horizons, 23, 194-200.

Manning, G. L., & Manning, M. (1984). What models of recreational reading make a difference. Reading World, 23, 375-380.

Morrow, L. M., & Weinstein, C. S. (1986). Encouraging voluntary reading: The impact of a literature program on children's use of library centers. Reading Research Quarterly, 21, 330-346.

Peak, J., & Dewalt, M. W. (1994). Reading achievement: Effects of computerized reading management and enrichment. ERS Spectrum, 12(1), 31-34.

Reutzel, D. R., & Hollingsworth, P. M. (1991). Reading comprehension skills: Testing the distinctiveness hypothesis. Reading Research and Instruction, 30, 32-46.

Summers, E. G., McClelland, J. V. (1982). A field-based evaluation of sustained silent reading (SSR) in intermediate grades. Alberta Journal of Educational Research, 28, 100-112.

Vollands, S. R., Topping, K. J., & Evans, R. M. (1999). Computerized self-assessment of reading comprehension with the Accelerated Reader: Action Research. Reading and Writing Quarterly, 15, 197-211.

HAPTER
4

Comprehension

COMPREHENSION

Executive Summary

Introduction

Comprehension is critically important to development of children's reading skills and therefore their ability to obtain an education. Indeed, reading comprehension has come to be viewed as the "essence of reading" (Durkin, 1993), essential not only to academic learning but to life-long learning. As the National Reading Panel (NRP) began its analysis of the extant research data on reading comprehension, three predominant themes emerged: (1) reading comprehension is a cognitive process that integrates complex skills and cannot be understood without examining the critical role of vocabulary learning and instruction and its development; (2) active interactive strategic processes are critically necessary to the development of reading comprehension; and (3) the preparation of teachers to best equip them to facilitate these complex processes is critical and intimately tied to the development of reading comprehension. With this as background, the NRP decided to organize its review and analysis of reading comprehension research in these three areas, and to address each in a subreport. This executive summary covers these three areas, and the format therefore differs slightly in organization from the other report executive summaries. Although the methodological issues pertinent to each of the three subareas are discussed in a common section, the results and discussion, as well as conclusions, implications for reading instruction, and directions for further research are combined under "Findings" for each of these three areas:

- Vocabulary Instruction
- Text Comprehension Instruction
- Teacher Preparation and Comprehension Strategies Instruction

This organization was adopted to accommodate the complexity of each of these subtopics. These reviews provide systematic evaluations and analyses of the research on these topics during the past 20 years.

Through the analyses, the Panel sought answers to this question: What methods are effective in teaching vocabulary and text comprehension and in preparing teachers to teach comprehension strategies?

Methodology

Database

To carry out scientific reviews, the NRP searched the research literature on vocabulary and text comprehension instruction from 1979 to the present. For vocabulary instruction, 47 studies met the NRP's scientific criteria. These studies included 73 grade-level samples, 53 of which were distributed from grades 3 to 8. For text comprehension instruction, 203 studies met the NRP's scientific criteria. These studies included 215 grade-level samples, 170 of which were distributed from grades 3 through 8. For preparation of teachers to teach text comprehension in naturalistic settings, the Panel intensively analyzed four relevant studies that appeared in the search of the text comprehension literature. These studies represent the only experimental attempts to prepare teachers to implement in naturalistic settings in the classroom instruction of proven text comprehension strategies that have evolved during the past 20 years. These studies also evaluated the effectiveness of the preparation on comprehension by the readers. The teacher preparation studies covered 53 classroom teachers from grades 2 to11.

Analysis

The Panel performed extensive analyses on the research studies identified in each subarea under review. The levels of analyses are briefly stated below:

Vocabulary Instruction

An exhaustive inquiry into recent research in vocabulary instruction techniques failed to elicit a numerically large database of studies that satisfied the NRP criteria for inclusion. Three meta-analyses included in the original search were analyzed separately from the instructional research studies. Although these analyses do not meet the formal criteria for inclusion in the analysis, they are relevant to the issues at hand. Consequently, they are included in the discussions of findings.

Text Comprehension Instruction

A study had to meet the following criteria to be included in the NRP review:

- Relevant to instruction of reading or comprehension. This criterion, in particular, excluded studies on comprehension instruction in reasoning and mathematics problem solving (Schoenfeld, 1985), physics (Larkin & Reif, 1976) and writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985).
- The study has to have been published in a scientific journal. A few exceptions are dissertations and conference proceedings that were reviewed in two meta-analyses by Rosenshine and his colleagues (Rosenshine & Meister, 1994; Rosenshine, Meister, & Chapman, 1996).
- The study had to have an experiment that involved at least one treatment and an appropriate control group or it had to have one or more quasi-experimental variables with variations that served as comparisons between treatments. The latter were rare.
- Insofar as could be determined, the participants or classrooms were randomly assigned to the treatment and control groups or were matched on initial measures of reading comprehension. This criterion was relaxed in a number of studies in which random assignment of classrooms was not carried out.

The Panel coded and entered the coded contents of the 205 studies that met these criteria into a database to identify the types of comprehension instruction that were reported as effective. First, the abstracts of the studies were examined and the type of instruction was coded, as were experimental treatments and controls (independent variables), grade and reading level of readers, instructor (teacher or experimenter), assessments (dependent variables), and type of text. The studies were then classified and grouped based upon the type of instruction used. A total of 16 distinct categories of instruction was identified.

Teacher Preparation and Comprehension Strategies Instruction

A study had to meet the following criteria to be included in the review:

- It focused on the preparation of teachers for conducting reading comprehension strategy instruction.
- The study had to have been published in a scientific journal.
- It was empirical.
- It was experimental, using random assignment, or quasi-experimental with initial matching on the basis of reading comprehension scores.
- The complete set of results of the study was reported.

Four studies met these criteria. The Panel developed a detailed outline of each of the selected studies, organized to permit comparison across studies. The Panel reviewed the research in reading comprehension instruction broadly and also selected certain specific topics for a deeper focus, that is, vocabulary and teacher preparation for teaching reading comprehension strategies. It should be noted that there are other relevant aspects of comprehension instruction, for example, instruction in listening comprehension and in writing, that the Panel did not address. In addition, the Panel did not focus on special populations such as children whose first language is not English and children with learning disabilities. It did not review the research evidence concerning special populations and thus cannot say that its conclusions are relevant to them.

Consistency With the Methodology of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. However, in most instances, the wide variations in methodologies and implementations required the Panel to qualify the use of the NRP criteria for evaluating single studies, multiple studies, and reviews of existing studies. These departures from the stated NRP criteria are briefly stated below and are discussed in greater detail in each of the reports.

Vocabulary Instruction

A formal meta-analysis was not possible. Inspection of the research studies that were included in the database revealed a heterogeneous set of methodologies, implementations, and conceptions of vocabulary instruction. The Panel found no research on vocabulary measurement that met the NRP criteria; therefore, a detailed review of implicit evidence is presented.

For the analysis of research on vocabulary instruction, there were recent meta-analyses that dealt with only the variables amenable to a meta-analysis. For the most part, the experimental research in vocabulary instruction involves many different variables and methodologies. It was deemed inappropriate to analyze such disparate studies as a group. For comprehension instruction, there were simply too many studies involving too many variables to allow for a simple meta-analysis. The decision was made to do the preliminary work necessary to organize comprehension instruction research for possible future analyses. For research on preparation of teachers to teach comprehension, there were only four studies, and a meta-analysis was not possible. The general analysis of teacher education and professional development found too few studies on too many variables to conduct a formal meta-analysis. Similarly, for computer technology and reading instruction, there were relatively few studies, most of which used variables that differed from each other.

Text Comprehension Instruction

A formal meta-analysis was not possible because even the studies identified in the same instructional category used widely varying sets of methodologies and implementations. Therefore, the Panel found few research studies that met all the NRP criteria; however, to the extent possible, NRP criteria were employed in the analyses. NRP criteria for evaluating existing reviews of research were used in the analyses of the two Rosenshine and colleagues meta-analyses.

Teacher Preparation and Comprehension Strategies Instruction

A formal meta-analysis was not possible because of the small number of studies identified. However, comprehensive summaries according to NRP guidelines for each of the four studies are included in the report.

Vocabulary Instruction: Findings

The importance of vocabulary knowledge has long been recognized. In 1925, the National Society for Studies in Education (NSSE) Yearbook (Whipple, 1925) noted: "Growth in reading power means, therefore, continuous enriching and enlarging of the reading vocabulary and increasing clarity of discrimination in appreciation of word values." (Davis, 1942, p. 76) presented evidence that comprehension comprises two "skills": Word knowledge or vocabulary and reasoning. Vocabulary occupies an important position in learning to read. As a learner begins to read, reading vocabulary encountered in texts is mapped onto the oral vocabulary the learner brings to the task. The reader learns to translate the (relatively) unfamiliar words in print into speech, with the expectation that the speech forms will be easier to comprehend. Benefits in understanding text by applying letter-sound correspondences to printed material come about only if the target word is in the learner's oral vocabulary. When the word is not in the learner's oral vocabulary, it will not be understood when it occurs in print. Vocabulary occupies an important middle ground in learning to read. Oral vocabulary is a key to learning to make the transition from oral to written forms. Reading vocabulary is crucial to the comprehension processes of a skilled reader.

Vocabulary Instruction Methods

Five main methods of teaching vocabulary were identified:

1. Explicit Instruction: Students are given definitions or other attributes of words to be learned.
2. Implicit Instruction: Students are exposed to words or given opportunities to do a great deal of reading.
3. Multimedia Methods: Vocabulary is taught by going beyond text to include other media such as graphic representations, hypertext, or American Sign Language that uses a haptic medium.
4. Capacity Methods: Practice is emphasized to increase capacity through making reading automatic.
5. Association Methods: Learners are encouraged to draw connections between what they do know and words they encounter that they do not know.

Results of Vocabulary Instruction

There are age and ability effects learning gains that occur from vocabulary instruction. These findings point to the importance of selecting age- and ability-appropriate methods.

1. Computer vocabulary instruction shows positive learning gains over traditional methods.
2. Vocabulary instruction leads to gains in comprehension.
3. Vocabulary can be learned incidentally in the context of storybook reading or from listening to the reading of others.
4. Repeated exposure to vocabulary items is important for learning gains. The best gains were made in instruction that extended beyond single class periods and involved multiple exposures in authentic contexts beyond the classroom.
5. Pre-instruction of vocabulary words prior to reading can facilitate both vocabulary acquisition and comprehension.
6. The restructuring of the text materials or procedures facilitates vocabulary acquisition and comprehension, for example, substituting easy for hard words.

Implications for Reading Instruction

These results indicate that

1. There is a need for direct instruction of vocabulary items required for a specific text.
2. Repetition and multiple exposure to vocabulary items are important. Students should be given items that will be likely to appear in many contexts.
3. Learning in rich contexts is valuable for vocabulary learning. Vocabulary words should be those that the learner will find useful in many contexts. When vocabulary items are derived from content learning materials, the learner will be better equipped to deal with specific reading matter in content areas.
4. Vocabulary tasks should be restructured as necessary. It is important to be certain that students fully understand what is asked of them in the context of reading, rather than focusing only on the words to be learned. Restructuring seems to be most effective for low-achieving or at-risk students.
5. Vocabulary learning is effective when it entails active engagement in learning tasks.
6. Computer technology can be used effectively to help teach vocabulary.
7. Vocabulary can be acquired through incidental learning. Much of a student's vocabulary will have to be learned in the course of doing things other than explicit vocabulary learning. Repetition, richness of context, and motivation may also add to the efficacy of incidental learning of vocabulary.
8. Dependence on a single vocabulary instruction method will not result in optimal learning. A variety of methods was used effectively with emphasis on multimedia aspects of learning, richness of context in which words are to be learned, and the number of exposures to words that learners receive.

Directions for Further Research

The need in vocabulary instruction research is great. Existing knowledge of vocabulary acquisition exceeds current knowledge of pedagogy. That is, a great deal is known about the ways in which vocabulary increases under highly controlled conditions, but much less is known about the ways in which such growth can be fostered in instructional contexts. There is a great need for the conduct of research on these topics in authentic school contexts, with real teachers, under real conditions.

1. What are the best ways to evaluate vocabulary size, use, acquisition, and retention? What is the role of standardized tests, what other measures should be used, and under what circumstances?
2. Given the preliminary findings that age and ability levels can affect the efficacy of various vocabulary instruction methods (Tomesen & Aarnoutse, 1998; Robbins & Ehri, 1994; Nicholson & Whyte, 1992; McGivern & Levin, 1983), what are the specific vocabulary instruction needs of students at different grade and ability levels?

3. What are the more general effects of vocabulary instruction across the grades?
4. Empirical support has been found for the facilitation of vocabulary learning with computers as ancillary aids and replacements of other technologies (Reinking & Rickman, 1990; Heise, 1991; Davidson et al., 1996). What is the optimal use of computer (and other) technologies in vocabulary instruction? What is the precise role of multimedia learning in vocabulary acquisition?
5. What is the precise role of multimedia learning in vocabulary instruction across the grades?
6. How should vocabulary be integrated in comprehension instruction for optimal benefit to the student?
7. What are the optimal combinations of the various methods of vocabulary instruction, including direct and indirect instruction, as well as different methods within these categories?
8. What sort of professional development is needed for teachers to become proficient in vocabulary instruction?

Text Comprehension Instruction: Findings

Comprehension is a complex process. There exist as many interpretations of comprehension as there are of reading. This may be so because comprehension is often viewed as "the essence of reading" (Durkin, 1993). Reading comprehension is further defined as "intentional thinking during which meaning is constructed through interactions between text and reader" (Durkin, 1993). According to this view, meaning resides in the intentional, problem-solving, thinking processes of the reader that occur during an interchange with a text. The content of meaning is influenced by the text and by the reader's prior knowledge and experience that are brought to bear on it. Reading comprehension is the construction of the meaning of a written text through a reciprocal interchange of ideas between the reader and the message in a particular text (Harris & Hodges, 1995, definition #2, p. 39).

The bulk of instruction of text comprehension research during the past 2 decades has been guided by the cognitive conceptualization of reading described above. In the cognitive research of the reading process, reading is purposeful and active. A reader reads a text to understand what is read and to put this understanding to use. A reader can read a text to learn, to find out information, or to be entertained. These various purposes of understanding require that the reader use knowledge of the world, including language and print. This knowledge enables the reader to make meanings of the text, to form memory representations of these meanings, and to use them to communicate information with others about what was read.

Although instruction on text comprehension has been a major research topic for more than 20 years, the explicit teaching of text comprehension before 1980 was done largely in content areas and not in the context of formal reading instruction. The idea behind explicit instruction of text comprehension was that comprehension could be improved by teaching students to use specific cognitive strategies or to reason strategically when they encountered barriers to comprehension in reading. The goal of such training was the achievement of competent and self-regulated reading.

Readers normally acquire strategies for active comprehension informally. Comprehension strategies are specific procedures that guide students to become aware of how well they are comprehending as they attempt to read and write. Explicit or formal instruction of these strategies is believed to lead to improvement in text understanding and information use. Instruction in comprehension strategies is carried out by a classroom teacher who demonstrates, models, or guides the reader in their acquisition and use. When these procedures are acquired, the reader becomes independent of the teacher. Using them, the reader can effectively interact with the text without assistance.

Comprehension Instruction Methods

Analyses of the 203 studies on instruction of text comprehension led to the identification of 16 different kinds of effective procedures. Of the 16 types of instruction, 8 offered a firm scientific basis for concluding that they improve comprehension. The eight kinds of instruction that appear to be effective and most promising for classroom instruction are

1. Comprehension monitoring in which the reader learns how to be aware or conscious of his or her understanding during reading and learns procedures to deal with problems in understanding as they arise.
2. Cooperative learning in which readers work together to learn strategies in the context of reading.
3. Graphic and semantic organizers that allow the reader to represent graphically (write or draw) the meanings and relationships of the ideas that underlie the words in the text.
4. Story structure from which the reader learns to ask and answer who, what, where, when, and why questions about the plot and, in some cases, maps out the time line, characters, and events in stories.
5. Question answering in which the reader answers questions posed by the teacher and is given feedback on the correctness.
6. Question generation in which the reader asks himself or herself what, when, where, why, what will happen, how, and who questions.
7. Summarization in which the reader attempts to identify and write the main or most important ideas that integrate or unite the other ideas or meanings of the text into a coherent whole.
8. Multiple-strategy teaching in which the reader uses several of the procedures in interaction with the teacher over the text. Multiple-strategy teaching is effective when the procedures are used flexibly and appropriately by the reader or the teacher in naturalistic contexts.

Results of Comprehension Instruction

With respect to the scientific basis of the instruction of text comprehension, the Panel concludes that comprehension instruction can effectively motivate and teach readers to learn and to use comprehension strategies that benefit the reader.

These comprehension strategies yield increases in measures of near transfer such as recall, question answering and generation, and summarization of texts. These comprehension strategies, when used in combination, show general gains on standardized comprehension tests. Teachers can learn to teach students to use comprehension strategies in natural learning situations. Furthermore, when teachers teach these strategies, their students learn them and improve their reading comprehension.

A common aspect of individual and multiple-strategy instruction is the active involvement of motivated readers who read more text as a result of the instruction. These motivational and reading practice effects may be important to the success of multiple-strategy instruction.

Multiple-strategy instruction that is flexible as to which strategies are used and when they are taught over the course of a reading session provides a natural basis for teachers and readers to interact over texts. The research literature developed from the study of isolated strategies to their use in combination to the preparation of teachers to teach them in interaction over texts with readers in naturalistic settings. The Panel regards this development as the most important finding of the Panel's review because it moves from the laboratory to the classroom and prepares teachers to teach strategies in ways that are effective and natural.

Implications for Reading Instruction

The empirical evidence reviewed favors the conclusion that teaching of a variety of reading comprehension strategies leads to increased learning of the strategies, to specific transfer of learning, to increased retention and understanding of new passages, and, in some cases, to general improvements in comprehension.

The important development of instruction of comprehension research is the study of teacher preparation for instruction of multiple, flexible strategies with readers in natural settings and content areas and the assessment of the effectiveness of this instruction by trained teachers on comprehension.

Directions for Further Research

The Panel's analysis of the research on instruction of text comprehension left a number of questions unanswered.

1. More information is needed on the effective ways to teach teachers how to use proven strategies for instruction in text comprehension. This information is crucial to situations in which teachers and readers interact over texts in real classroom contexts.
2. Some evidence was reviewed that indicated that instruction in comprehension in content areas benefits readers in terms of achievement in social studies. However, it is not clear whether instruction of comprehension strategies leads to learning skills that improve performance in content areas of instruction. If so, then it might be efficient to teach reading comprehension as a learning skill in content areas.
3. Instruction of comprehension has been successful over the 3rd to 6th grade range. A next step will be to determine whether certain strategies are more appropriate for certain ages and abilities, what reader characteristics influence successful instruction of reading comprehension, and which strategies, in combination, are best for younger readers, poor or below-average readers, or for learning-disabled and dyslexic readers.
4. It will be important to know whether successful instruction generalizes across different text genres (e.g., narrative and expository) and texts from different subject content areas. The NRP review of the research indicated that little or no attention was given to the kinds of text that were used and that there was little available information on the difficulty level of texts.
5. It will also be important to determine what teacher characteristics influence successful instruction of reading comprehension and what the most effective ways are to train teachers, both preservice and inservice.
6. Criteria of internal and external validity should be considered in the design of future research, to address problems that were noted in prior studies. Specifically, these issues were random assignment of students to treatments and control conditions; exposure of experimental and control participants to the same training materials; provision of information about the amount of time spent on dependent variable tasks; the study of fidelity of treatment and analyzing teacher and reader performance during instruction; use of appropriate units (individual, group, classroom) in analyses; and assessment of either long-term effects or generalization of the strategies to other tasks and materials.

Teacher Preparation and Comprehension Strategies Instruction: Findings

The preparation of teachers to deliver comprehension strategy instruction is important to the success of teaching reading comprehension. As indicated by the Panel's review of text comprehension, reading comprehension can be improved by teaching students to use specific cognitive strategies or to reason strategically when they encounter barriers to comprehension when reading. The goal of such training is the achievement of competent and self-regulated reading. The research on comprehension strategies has evolved dramatically over the course of the last 2 decades. At first, investigators focused on teaching students one strategy at a time. A wide variety of strategies was studied, including imagery, question generating, prediction, and a host of others. In later studies, several strategies were taught in combination. However, implementation in the context of the actual classroom of this promising approach to comprehension has been problematic. Acquiring and practicing strategies in isolation and then attempting to provide transfer opportunities during the reading of text is not the kind of instruction that is required in naturalistic contexts. Proficient reading involves a constant, ongoing adaptation of many cognitive processes.

Thus, teachers must be skillful in their instruction and must respond flexibly and opportunistically to students' needs for instructive feedback as they read. To be able to do this, teachers must themselves have a firm grasp not only of the strategies that they are teaching the children but also of instructional strategies that they can employ to achieve their goal. Many teachers find this type of teaching a challenge, most likely because they have not been trained to do such teaching. The focus of the review was on four recent and promising studies that addressed the need for specific teacher preparation in the implementation of strategy instruction in naturalistic classroom contexts. In these four studies, teachers were trained to teach strategies, and the focus was on the effectiveness of that training on students' reading. It is not surprising that only a few relevant studies have been done on this topic. Interest in the topic is rather new, and preparing teachers to deliver effective strategy instruction is a lengthy and complex process.

Methods of Teacher Preparation

There have been two major approaches to comprehension strategy instruction in the classroom: direct explanation (DE) and transactional.

The DE approach was designed to improve on the approach in which students are taught to use one or several strategies as described and reviewed in the previous section on text comprehension instruction. This kind of instruction did not attempt to provide students with an understanding of the reasoning and mental processes involved in reading strategically. Therefore, Gerald Duffy and Laura Roehler developed the DE approach. In this approach, teachers do not teach individual strategies but focus instead on helping students to (1) view reading as a problem-solving task that necessitates the use of strategic thinking and (2) learn to think strategically about solving reading comprehension problems. The focus is on developing teachers' ability to explain the reasoning and mental processes involved in successful reading comprehension in an explicit manner, hence the use of the term "direct explanation." The implementation of DE requires specific and intensive teacher training on how to teach the traditional reading comprehension skills found in basal readers as strategies, for example, to teach students the skill of how to find the main idea by casting it as a problem-solving task and reasoning about it strategically. The transactional strategy instruction (TSI) approach includes the same key elements as the direct explanation approach, but it takes a somewhat different view of the role of the teacher in strategy instruction. The TSI approach focuses on the ability of teachers to facilitate discussions in which students (1) collaborate to form joint interpretations of text and (2) explicitly discuss the mental processes and cognitive strategies that are involved in comprehension. In other words, the emphasis is on the interactive exchange among learners in the classroom, hence use of the term "transactional."

In both approaches, teachers explain specific strategies to students and model the reasoning associated with their use. Both approaches include the use of systematic practice of new skills, as well as scaffold support, in which teachers gradually withdraw the amount of assistance they offer to students. The different emphases of the two approaches (explanation vs. discussion) result in differences in the level of collaboration among students.

Results of Teacher Preparation

All four studies showed that teachers can be taught to be effective in teaching comprehension to their students in naturalistic reading contexts. These studies indicate that teaching teachers to use comprehension instruction methods leads to students' awareness of strategies and use of strategies, which can in turn lead to improved reading.

Implications for Reading Instruction

Teachers need training to become effective in explaining fully what it is that they are teaching (what to do, why, how, and when), modeling their own thinking processes for their students, encouraging students to ask questions and discuss possible answers and problem solutions among themselves, and keeping students engaged in their reading by providing tasks that demand active involvement.

There should be greater emphasis in teacher education on the teaching of reading comprehension. Such instruction should begin during preservice training, and it should be extensive, especially with respect to preparing teachers to teach comprehension strategies.

Directions for Further Research

It will be important to determine how well teachers maintain their effectiveness in the classroom after they have had preparation in teaching reading comprehension. Thus, several research questions remain to be investigated.

1. Which components of a successful teacher preparation program are the effective ones? These could include characteristics of the teacher preparation program itself, such as its focus or

intensity, as well as characteristics of the instruction delivered to the students, such as the amount of instruction provided, the particular strategies taught, and the amount of collaborative discussion involved.

2. Can instruction in reading comprehension strategies be successfully implemented and incorporated into content area instruction?
3. How should the effectiveness of strategy instruction be assessed, especially with respect to reading achievement and subject matter proficiency, but also with respect to student interest and teacher satisfaction?
4. Comprehension instruction research has been limited to grades 3 through 8. It will be important to examine comprehension instruction in the primary grades when children are mastering phonics and word recognition and developing reading fluency.
5. How should the effectiveness of strategy instruction be assessed, for example, by reading achievement or subject matter achievement?
6. Should the teaching of reading comprehension begin during preservice or inservice training or both, and how extensive should it be?

Conclusions

Vocabulary is one of the most important areas within comprehension and should not be neglected. The NRP found a variety of methods by which readers acquire vocabulary through explicit instruction and improve their comprehension of what they read. The Panel also found that although there has been considerable success in teaching a variety of effective text comprehension strategies that lead to improved text comprehension, the most promising lines of research within the reading comprehension strategies area focused on teacher preparation to teach comprehension. Teachers can be helped by intensive preparation in strategy instruction, and this preparation leads to improvement in the performance of their students.

COMPREHENSION

Introduction

> Learning to read is one of the most important things children accomplish in elementary school because it is the foundation for most of their future academic endeavors. From the middle elementary years through the rest of their lives as students, children spend much of their time reading and learning information presented in text. The activity of reading to learn requires students to comprehend and recall the main ideas or themes presented in . . . text. (Stevens et al., 1991, p. 8).

Inherent in the view expressed by Stevens et al. (1991) in the preceding quotation is the critical importance of reading comprehension and the developmental view that children must first learn how to recognize and relate print to oral language knowledge and to make this recognition automatic through practice. Indeed, comprehension has come to be viewed as the "essence of reading" (Durkin, 1993), essential not only to academic learning but to lifelong learning as well. Despite the critical and fundamental importance of reading comprehension, comprehension as a cognitive process began to receive scientific attention only in the past 30 years.

As the National Reading Panel initiated its analysis of the extant research data on reading comprehension, three major themes emerged. First, reading comprehension as a process and an amalgam of complex skills cannot be understood without examining the critical role and importance of vocabulary and vocabulary instruction. Second, in contrast to earlier speculation that reading comprehension was a passive process, more recent data clearly indicate that robust comprehension is dependent on active and thoughtful interaction between the text and the reader. Therefore, the development and investigation of reading comprehension strategies now occupy a central role in teaching readers how to maximize their understanding of what is read and in explaining the different strategies that enable readers to optimally understand text. Third, given the importance of comprehension strategies in fully understanding text, it is critical to know how teachers can best be prepared to teach their students specific strategies that will facilitate this understanding.

With this as background, this section of the NRP report is organized in three subsections. The first section examines specific evidence on the effects of vocabulary instruction on reading achievement, particularly text comprehension. In the second section, data relevant to the effects of different types of instruction to facilitate the comprehension of text are analyzed and interpreted. In the third section, available data on the preparation of teachers to teach reading comprehension strategies are reviewed and interpreted. In addition, a comprehensive description of the review methodology employed in the evaluation of studies relevant to vocabulary instruction, text comprehension instruction, and teacher preparation is provided in Appendix A of this report.

In contrast to the sections in the NRP report that address phonemic awareness, phonics, and fluency, studies undertaken with young children at risk for reading failure and older disabled readers were not reviewed and analyzed as part of the Comprehension report. Time and resource limitations prevented the Panel from examining this subtopic in an optimal manner. Thus, it was decided to examine only literature that pertained to the normal reading process.

The reader will also note that the organization of this section of the report differs from the structure employed in adjacent sections in that each subsection (Vocabulary Instruction, Text Comprehension Instruction, Teacher Preparation and Comprehension Strategies Instruction) provides a topic-specific introduction, an overview of methodology, results and discussion of the findings, implications for reading instruction, and directions for future research. This organization was adopted to accommodate the complexity of each of these subtopics and differences in the review and data analytic procedures employed for each of the topics.

HAPTER 4

Comprehension

Part I

Vocabulary Instruction

COMPREHENSION I

Vocabulary Instruction

Introduction

The importance of vocabulary in reading achievement has been recognized for more than half a century. As early as 1925, in the National Society for Studies in Education (NSSE) Yearbook, this quotation appears:

> Growth in reading power means, therefore, continuous enriching and enlarging of the reading vocabulary and increasing clarity of discrimination in appreciation of word values (Whipple, 1925, p. 76).

Even today, evidence of the importance of vocabulary is usually attributed to Davis (1942), who presented evidence that comprehension comprised two "skills": word knowledge or vocabulary and reasoning in reading. The Panel reflects this position with the inclusion of the current analysis of research on vocabulary instruction with the other comprehension research analyses. Since Davis' work, there have been questions regarding the "skills" perspective, but the finding that vocabulary is strongly related to comprehension seems unchallenged.

Given the prominence of vocabulary in the reading process, the comprehension subgroup determined that vocabulary instruction merited a specific review. Therefore, the purpose of this report was to examine the scientific evidence on the effect of vocabulary instruction on reading achievement. This was done is two stages: first examining the literature on vocabulary instruction and, second, the literature on the measurement of vocabulary.

Vocabulary Instruction

Vocabulary occupies an important position in learning to read. As a learner begins to read, reading vocabulary encountered in texts is mapped onto the oral vocabulary the learner brings to the task. That is, the reader is taught to translate the (relatively) unfamiliar words in print into speech, with the expectation that the speech forms will be easier to comprehend. A benefit in understanding text by applying letter-sound correspondences to printed material only comes about if the resultant oral representation is a known word in the learner's oral vocabulary. If the resultant oral vocabulary item is not in the learner's vocabulary, it will not be better understood than it was in print. Thus, vocabulary seems to occupy an important middle ground in learning to read. Oral vocabulary is a key to learning to make the transition from oral to written forms, whereas reading vocabulary is crucial to the comprehension processes of a skilled reader.

Despite the clear importance of vocabulary, recent research has focused more on overall comprehension than on vocabulary. This appears to be a function of the more inclusive nature of many contemporary comprehension methods, which seem to incorporate at least some vocabulary instruction. Even in traditional methods of teaching reading, lesson formats always include vocabulary instruction.

Many studies have shown that reading ability and vocabulary size are related, but the *causal* link between increasing vocabulary and an increase in comprehension has not been demonstrated. That is, it has been difficult to demonstrate that teaching vocabulary improves reading ability.

Why this should be so difficult is sometimes obscured by the imprecise nature of the definitions of vocabulary and comprehension. Both vocabulary and comprehension involve the meaning of the text, albeit at different levels. Vocabulary is generally tied closely to individual words while comprehension is more often thought of in much larger units. To get to the comprehension of larger units requires the requisite processing of the words. Precisely separating the two processes is difficult, if not impossible.

Measurement of Vocabulary

Even the measurement of vocabulary is fraught with difficulties. Researchers distinguish between many different "vocabularies." Receptive vocabulary is the vocabulary that we can understand when it is presented to us in text or as we listen to others speak, while productive vocabulary is that vocabulary we use in writing or when speaking to others. It is generally believed that receptive vocabulary is much larger than productive vocabulary since we often recognize words that we would rarely use. Vocabulary is also subcategorized as oral vs. reading vocabulary, where oral refers to words that are recognized in speaking or listening while reading vocabulary refers to words that are used or recognized in print. Sight vocabulary is a subset of reading vocabulary that does not require explicit word recognition processing. Conclusions about some of these different types of vocabularies often do not apply to all; what may be true for one may or may not be true for another.

At a conceptual level, vocabulary can be measured in many ways. One major distinction in the measurement of vocabulary parallels the receptive/productive distinction. Vocabulary that is *recognized* by an individual is often different from vocabulary that is *produced*. Another distinction is made between reading vocabulary and writing vocabulary—the vocabularies that are available to the reader or writer—and between speaking and listening vocabularies. Still another type of vocabulary is often referred to as sight vocabulary—those words that can be identified without explicit decoding during reading.

Because there are so many definitions of vocabulary, the format for assessing or evaluating vocabulary is an important variable in both practice and research. One way of assessing recognition vocabulary is to have the learner select a definition for a word from a list of alternatives. Conversely, the task *could* be to select a word for the definition. In many cases, such as standardized tests, this method is used as a means of obtaining efficiency in testing. A second method of assessing vocabulary is by having the learner generate a definition for a word. Because this method requires a judgment about the response, it is often deemed less efficient than a recognition method. Most often, recognition vocabulary is measurably larger than productive vocabulary.

Another difficulty with the measurement of vocabulary is that we can only ask a learner for a relatively small number of words. Those words must be representative of a larger pool of vocabulary items. In short, we can never know *exactly* how large a vocabulary an individual has. Instead, we often measure only specific vocabulary items that we want the individual to know, for example, in the context of a reading or a science lesson. Standardized tests attempt to deal with this by selecting words that differ widely in their familiarity. Persons who can correctly identify unfamiliar words are assumed to have larger vocabularies. The more unfamiliar words that can be identified, the larger the vocabulary. However, these are estimates, rather than precise measurements. Furthermore, the definition of "familiar" or "known" words is difficult to pin down outside of a specific context. What does it mean if a learner "almost" knows a word? The assessment of such a circumstance has no objective answer.

Finally, evaluation of vocabulary knowledge is measured either by standardized tests or by informal, experimenter- or teacher-generated tests on one dimension and by receptive vs. productive techniques on another dimension.

Methodology

Database

A search using Endnote 3.0 connected to the ERIC online database with a Z39.50 connection was initiated. Using the term "vocabulary" alone (in any field) yielded 18,819 citations. A search using "vocabulary" and "instruction" and "reading" and "research" and "method" yielded 141 citations. A similar search undertaken using the PsycINFO database yielded a total of 56 nonoverlapping citations. The 197 citations were downloaded into an Endnote library for further analysis. From this set, citations were removed if they were not reports of research, did not report experimental or quasi-experimental studies, dealt with foreign languages or non-English-speaking groups, or dealt exclusively with learning disabled or other special populations, including second-language learners.

There are many studies that describe aspects of vocabulary without specifically addressing the questions of how vocabulary instruction is conducted. The Panel does recognize the importance of many of these studies in designing vocabulary instruction, but the Panel did not analyze these studies unless they contained at least some experimental work on instructional methods.

Additional bibliographic searching was conducted, guided by three meta-analyses (Stahl & Fairbanks, 1986; Klesius & Searls, 1990; Fukkink & de Glopper, 1998) and two reviews of the literature on vocabulary instruction research (Nagy & Scott, in press; Blachowicz & Fisher, in press). These procedures yielded a total of 50 studies that were candidates for further analysis. The studies were coded in a Filemaker 4.0 database, using the categories established by the NRP.

As the Panel analyzed the studies in the database, the Panel found no research that met the NRP criteria that explicitly addressed the issues of measuring vocabulary. This is clearly a gap in our knowledge and a research need.

Analysis

An exhaustive inquiry into recent research in vocabulary instruction techniques failed to elicit a numerically large database of studies that satisfied the NRP criteria for inclusion. Although the small size of the database of experimental research might temper some of the conclusions from the data, important and interesting trends do appear in the body of available studies. Following is a discussion of some salient observations from the extant data set, as well as some preliminary analyses of trends and important findings.

Three meta-analyses included in the original search were analyzed separately from the instructional research studies. Although these analyses do not meet the formal criteria for inclusion in the analysis, they are relevant to the issues at hand. Consequently, they are included in the discussions of findings.

Consistency With the Methodology of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. A formal meta-analysis was not possible. Inspection of the research studies that were included in the database revealed a heterogeneous set of methodologies, implementations, and conceptions of vocabulary instruction. As noted, the Panel found no research on vocabulary measurement that met the NRP criteria; therefore, implicit evidence is presented below on this issue.

Results

Summary and Preliminary Taxonomy of Instruction Methods

Because so many of these studies examined involve unique instructional programs, it was deemed appropriate to provide a summary of the methods used to study vocabulary. Table 1 in Appendix A lists the methods, a description of the basic techniques, and some sample citations for the method.

Because there were so many different methods represented in the database, a scheme for categorizing the methods was attempted. There are so many dimensions on which vocabulary instruction can be categorized that each implementation often appears to be unique. This seems to be the case for two reasons. First, there are typically so few vocabulary studies that each seems to distinguish itself from others by its differences from rather than its similarities to other methods. The second reason is that the similarities between methods have not been systematically organized at the conceptual level. The following scheme is an attempt to produce a simplified taxonomy of methods for vocabulary instruction.

Explicit Instruction

In explicit instruction, students are given definitions or other attributes of words to be learned. They are often given specific algorithms for determining meanings of words, or they are given external cues to connect the words with meaning. A common example of this technique is the pre-teaching of vocabulary prior to reading a selection. Other common methods of explicit instruction involve the analysis of word roots or affixes.

Indirect Instruction

In indirect instruction, students are exposed to words or given opportunities to do a great deal of reading. It is assumed that students will infer any definitions they do not have. At least one version of the implicit methods simply suggests that students should be encouraged to do wide reading to increase vocabulary.

Multimedia Methods

In these methods, vocabulary is taught by going beyond text to include other media. Semantic mapping and graphic representations of word attributes are among these methods (Margosein, Pascarella, & Pflaum, 1982; Levin, Johnson, Pittelman, Levin, Shriberg, Toms-Bronowski, & Hayes, 1984.) Newer developments like hypertext go beyond the single medium of text in attempts to enhance vocabulary learning. American Sign Language (Daniels, 1994, 1996) has been used to increase vocabulary, capitalizing on encoding in a haptic medium.

Capacity Methods

At least a few methods attempt to reduce the cognitive capacity devoted to other reading activities by practicing them to make them more nearly automatic. These methods assume that the additional capacity freed up can be used for vocabulary learning. These methods work to allow the student to concentrate on meaning of words rather than their orthographic or oral representations.

Association Methods

In this category of methods, learners are encouraged to draw connections between what they do know and words they encounter that they do not know. Sometimes these associations are semantic or contextual. At other times, they are based on imagery students invoke in learning the words.

Conclusion About the Taxonomy

Although the taxonomic scheme developed above describe the research at a general level, the Panel found that the differences between studies within the taxonomy were too great to be useful. In addition, many of the studies seemed to combine elements that would place them in one or more categories when the actual methods were developed. Consequently, although the Panel thinks it is important to think about vocabulary along these dimensions, the taxonomy was only used in a conceptual manner in subsequent analyses of the vocabulary instruction studies.

Analysis of the Research Studies

In the following analysis, the reading instruction database was reviewed for trends across studies, accounting for the great diversity in methods and the relatively small number of studies. The fact that the same studies are represented in more than one finding testifies to the complex nature of the instruction represented by many methods. For each of the trends, representative examples of studies are included with brief sketches of the findings.

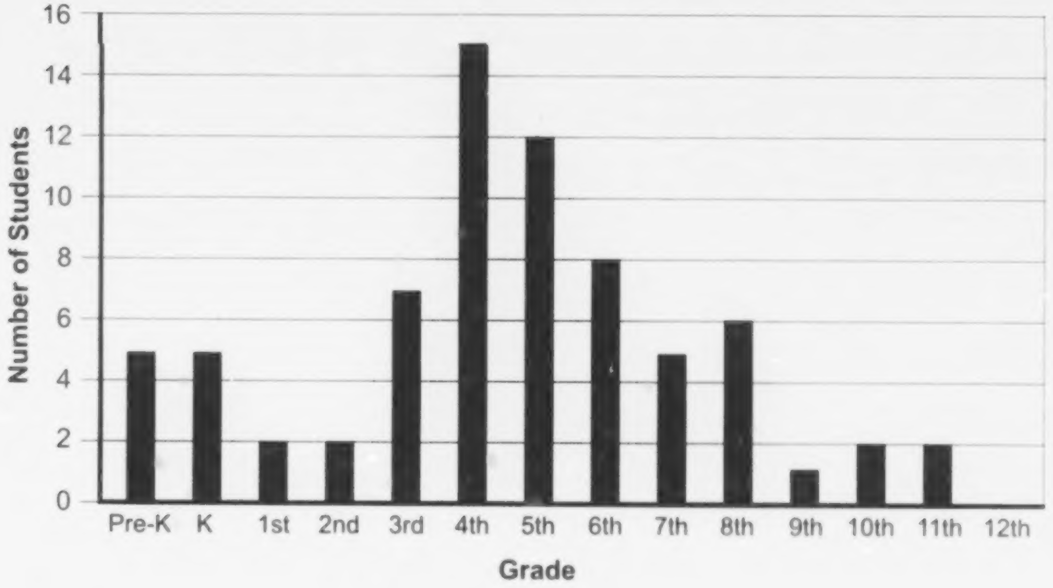
Age and Ability Effects on Vocabulary Learning

The distribution of research studies in vocabulary instruction as a function of grade level is shown in Figure 1 on the next page. What is most striking in these data is the fact that there are relatively few studies outside the range of 3rd to 8th grade. For the 50 studies categorized, there were 73 different grade samples because some studies used more than one grade level. Of these 73 grade samples, 53 were grades 3 to 8, with relatively little research on vocabulary instruction in the early grades. One possible explanation is that there is less emphasis on methods in the early grades. Another is that teaching of vocabulary is often not separate from other instruction in the early grades. As students begin to read content material they may need to learn vocabulary specific to the material, giving rise to the instructional need for vocabulary learning. Another possibility is that much of early reading is, at least theoretically, done with texts that do not exceed the vocabularies of most early readers. In this event, there would be little need for vocabulary instruction.

Despite the restricted range of studies, one trend in the database suggests that various ability levels and age differences can significantly affect learning gains from vocabulary instruction methods. The studies underscore the need to consider carefully the different impacts that various vocabulary instruction techniques can have for students of different ages and abilities, and, accordingly, the importance of selecting appropriate methods.

- Senechal and Cornell (1993) found that a single book reading was enough to significantly improve children's new expressive vocabulary of ten target words in the stories, and that after 1 week, the 5-year-olds remembered more than 4-year-olds.

Figure 1. NIH Vocabulary Studies: Distribution of Grades Studied in Research (N = 72 Grade samples in 47 studies)



- Meyerson, Ford, and Jones (1991) found that 5th graders were more likely than 3rd graders to assign science vocabulary into conceptual groupings.
- Tomesen and Aarnoutse (1998) studied reciprocal teaching and direct instruction in deriving word meanings from context as provided to 4th graders; the instruction was more helpful for poor readers rather than average readers.
- Robbins and Ehri (1994) found that storybook readings helped teach children meanings of unfamiliar words; those with larger entering vocabularies learned more words.
- Nicholson and Whyte (1992) explored how 8- to 10-year-old students learned vocabulary from incidental exposure (listening to stories). The largest effects were for high-ability students. They propose that low-ability and average students should do more independent reading with a dictionary than listening to stories.
- McGivern and Levin (1983) reported positive effects for the keyword method, with greater effects for low- than for high-ability students; the low-ability students had more difficulty in operationalizing dual components of the task.

Computer Use for Vocabulary Instruction

A small but clear trend in recent years shows computer technology making inroads in literacy and literacy instruction. Four studies that employ computers for vocabulary instruction appear in the database. These studies show learning gains with computer use as compared to traditional methods or when computers are used as an ancillary aid.

- Heller, Sturner, Funk, and Feezor (1993) examined the issue of cognitive demands of technology for preschool learners, by studying the effect of different input devices (touch screen vs. keyboard) on vocabulary identification. They concluded that the greater cognitive demands of keyboard use disrupted the children's ability to process the limited acoustic information available in speech.
- Reinking and Rickman (1990) found that 6th grade students receiving computer instruction of difficult text words with electronic text scored higher on vocabulary measures than students reading printed pages with dictionaries or glossaries.
- Heise, Papelweis, and Tanner (1991) compared 3rd and 6th through 8th grade students in conditions with computer-assisted and conventional direct instruction; the trend was for improved

performance with computer assistance, although the difference was not statistically significant.

- Davidson, Elcock, and Noyes (1996) used a computer that gave speech prompts when the learner requested them; 5- to 7-year-old students improved on three measures of vocabulary with these prompts.

Vocabulary Instruction Effects on Comprehension

In this category are studies that attempt to map the causal relationships between vocabulary and comprehension. The following studies underscore the notion that comprehension gains and improvement on semantic tasks are results of vocabulary learning. Although all of these studies focus on vocabulary, they also typify the heterogeneity among definitions and implementations of vocabulary instruction.

- Beck, Perfetti, and McKeown (1982) demonstrated that 4th graders receiving vocabulary instruction performed better on semantic tasks than those who did not receive instruction.
- McKeown, Beck, Omanson, and Perfetti (1983) also found that vocabulary instruction had a strong relation to text comprehension for 4th grade students.
- Wixson (1986) examined teaching the concept vs. dictionary definitions and showed that pre-teaching vocabulary words for understanding was effective, although the precise effects were unclear because of interaction with story.
- Carney, Anderson, Blackburn, and Blessings (1984) found that for 5th grade students, pre-teaching vocabulary words had a significant effect on retention and acquisition of social studies content.
- Kameenui, Carnine, and Freschi (1982) found that substitution of easy for hard vocabulary words, inclusion of redundant information, and instruction on difficult words facilitated comprehension.
- Stahl and Fairbanks (1986) conducted a meta-analysis and concluded that vocabulary instruction was an important component for comprehension. The best instructional techniques were mixes of definitional and contextual programs; the keyword method produced some significant gains in recall. Repeated exposures to words were also found to be effective.
- Medo and Ryder (1993) found that text-specific vocabulary instruction prior to reading expository texts helped 8th grade students to make causal connections and that this method benefited both average and high-ability students.

However, one study found that vocabulary instruction did not transfer to general reading comprehension. Tomesen and Aarnoutse (1998) conducted vocabulary instruction in the context of reciprocal teaching for 4th grade students. They used direct instruction in deriving word meanings from context and found it to be more helpful for poor readers than for average readers, but they reported a lack of transfer to general reading comprehension.

Keyword Method

In the database, some positive findings with the keyword method research indicate that this method may significantly augment recall, and may be more helpful than many other vocabulary instruction methods. One study found that the keyword method interacts with student ability levels, and that low-ability students had considerably more difficulty with certain keyword methods than high-ability students. However, another study reported that the initial keyword gains were temporary, fading out within a week.

- Levin and colleagues (1984) noted gains for 4th and 5th grade students with the keyword method as compared to semantic and contextual analysis methods in the short term. However, the advantage had faded in the 1-week-delayed test.
- Levin, McCormick, Miller, and Berry (1982) found that 4th grade students outperformed controls in vocabulary acquisition with the keyword method as compared to the picture context, control and experiential context conditions.
- Levin, Levin, Glasman, and Nordwall (1992) found strong effects for 3rd, 4th, 7th, and 8th grade students when comparing the keyword method to free study and science context vocabulary methods.
- McGivern and Levin (1983) found that 5th grade students showed positive effects of the keyword method. However, there was more of a difference for low-ability students than for high-ability students, although low-ability students had more difficulty in operationalizing the components of the task.

Indirect Learning Effects

Because of the rapid rate at which vocabulary is acquired, it has always been assumed that much vocabulary was learned incidentally. One instantiation of this method is found in vocabulary learning in the context of storybook reading. Recent research studies in the area suggest that indirect learning can definitely occur, and that vocabulary can be acquired through incidental exposure. In addition, one particular study (Schwanenflugel, Stahl, & McFall, 1997) is important because it looks beyond the issue of whether word acquisition occurs from reading, examining the characteristics of words and texts that were most amenable to vocabulary acquisition from stories. In this study of 4th grade students, researchers found that non-noun words (adverbs, verbs, and adjectives) were easier to learn than nouns and that words with high imageability were easier to learn from the stories.

- Robbins and Ehri (1994) demonstrated that storybook readings helped teach children meanings of unfamiliar words. However, those with larger entering vocabulary learn more words.
- Leung (1992) studied kindergartners and 1st grade students, finding that the frequency of a target word in stories influenced the occurrence of the word in the child's retellings and that read-aloud events seemed to help children to learn new words by incidental learning.
- Senechal and Cornell (1993) found that for 4- to 5-year-old children, one single book reading was enough to significantly improve new expressive vocabulary of ten target words in the stories. In a delayed transfer test after 1 week, 5-year-old children remembered more than 4-year-olds.
- Nicholson and Whyte (1992) explored student vocabulary learning through incidental exposure by having children 8 to 10 years old listen to stories; the largest effects were for high-ability students. They proposed that low-ability and average-ability students do more independent reading with a dictionary than listening to stories.
- Stewart, Gonzalez et al. (1997) examined acquisition of sight-reading vocabulary learned incidentally during articulation training and found that this learning generalized beyond printed words on cards to words on a list.
- Stahl, Richek, and Vandevier (1991) evaluated the indirect learning of vocabulary words among 6th grade students designated as less able readers and found that the students were able to learn a significant number of vocabulary words from listening to orally presented passages.

Two studies revealed great detail about the actual process of vocabulary learning by examining the characteristics of words that were most conducive to vocabulary acquisition. Schwanenflugel, Stahl, and McFalls (1997) found that among their 4th grade sample, certain word characteristics had a significant impact on vocabulary learned from reading stories. In particular, non-noun words (verbs, adverbs, and adjectives) were learned better than nouns, and concrete words (high in imageability) were learned more readily than less easily imageable words. The authors conclude that the characteristics of vocabulary words are more important variables in the learning of vocabulary words from stories than are text features (word repetitions, contextual support, etc.). Another study, McFalls, Schwanenflugel, and Stahl (1996), examined the impact of semantic variables related to concreteness on the development of reading vocabulary among a predominantly African American and low SES 2nd grade sample. They found that the children read abstract words with less accuracy than concrete words on tasks of recognition and reading accuracy and that the concreteness of the words determined whether children were able to remember them and to learn to read them more easily.

The nature of the interaction (emphasizing active participation) during storybook readings may also have an impact on learning. Three studies found that student-initiated talk or active participation was important.

- Dickinson and Smith (1994) examined storybook readings for preschoolers and the effects of teacher talk on vocabulary acquisition and concluded that the amount of child-initiated analytic talk was important for vocabulary gains.
- Senechal (1997) found that for pre-kindergarten children, repeated readings of a story created greater performance gains in vocabulary. Students learned more from answering questions during readings than they did when simply listening to the narrative.

- Drevno, Kimball, Possi, Heward, Gardner, and Barbetta (1994) examined the effects of active student response (ASR) error correction on the learning of science vocabulary for a small group of elementary students. In the ASR condition, when a student made an error, the teacher modeled the correct definition and the student repeated it, but in the no response (NR) condition, students would not repeat the definition. ASR was found to be superior to the NR error-correction condition on all the dependent variables.

Vocabulary Gains From Repeated, Multiple Exposures

One trend that was strongly reflected in the database was that high frequency and multiple, repeated exposures to vocabulary material are important for learning gains. In accordance with this finding, a trend was also noted that extended and rich instruction of vocabulary (applying words to multiple contexts, etc.) was superior to less comprehensive methods. The following studies share this finding:

- Senechal (1997) found that for pre-kindergarten children, repeated readings of a story were associated with greater performance gains in vocabulary.
- Leung (1992) studied kindergarten and 1st grade students, finding that the frequency of a target word in stories influenced occurrence of the word in a child's retellings.
- Daniels, M. (1994) showed that pre-K students who learned American Sign Language (ASL) did significantly better than controls on the Peabody Picture Vocabulary Test (PPVT). In a 1996 study, Daniels also found that kindergarten students who learned ASL did significantly better on language development and vocabulary growth measures of the PPVT than those who had not learned ASL.

Effect of Rich Contexts on Vocabulary Growth

- McKeown, Beck, Omanson, and Pople (1985) found that 4th graders performed well with instruction that extended beyond single class periods and involved multiple exposures in authentic contexts. The instruction added activities to extend use of learned words beyond the classroom and high-frequency encounters with words.
- Kameenui, Carnine, and Freschi (1982) found that providing redundant information facilitated comprehension and that instruction on difficult vocabulary words also helped vocabulary learning in grades 4 through 6.
- Dole, Sloan, and Trathen (1995) worked with 10th grade students on an "alternative" vocabulary treatment condition: teach students how to select relevant words, learn the words on a deep level, and discuss them. These students outscored students taught with the traditional conditions in which students did not learn this criterion or discuss the words in context.

Pre-instruction of Vocabulary Words

It has been a given for reading instruction in almost every formal lesson format that vocabulary instruction will occupy a central part of the lesson, typically prior to reading. This pre-instruction has often been justified on the basis of making the passage easier to comprehend by reducing the cognitive load during subsequent reading. In fact, a few studies suggest that pre-instruction of vocabulary words facilitates both vocabulary acquisition and comprehension.

- Brett, Rothlein, and Hurley (1996) found that 4th grade students who were given pre-instruction of target words in the story had greater vocabulary gains than the children in the non-instructional control group.
- Wixson (1986) pre-taught vocabulary words to grade students. Although there were some gains in understanding, the instructional treatment (concept vs. dictionary) effects were unclear because of interaction with story.
- Carney, Anderson, Blackburn, and Blessing (1984) also pre-taught vocabulary to 5th grade students; the treatment had a significant effect on retention and acquisition of social studies content.

Restructuring the Task

One emergent trend in the database is the restructuring of the task (materials or procedures) in various ways to facilitate vocabulary acquisition and comprehension. A way of doing this is to alter the passage, such as substituting easy for hard words. Another is clarifying the task of learning vocabulary definitions for students, such as teaching what components make a good definition, and selecting relevant words. Group-assisted reading in student dyads also yielded significant vocabulary gains over the comparison, unassisted group. Although the diversity among these studies is a salient feature, the following studies did found positive results with a wide range of task alterations:

- Kameenui, Carnine, and Freschi (1982) found that providing redundant information facilitated comprehension and that instruction on difficult vocabulary words also helped vocabulary learning in grades 4 through 6.
- Gordon, Schumm, Coffland, and Doucette (1992) revised text versions to help define vocabulary words for 5th grade students. Using these revised texts helped students understand passages better.
- Schwartz and Raphael (1985) clarified the task of defining a word for 4th and 5th grade students, giving them the components of a definition; this increased students' independent vocabulary acquisition.
- Scott and Nagy (1997) evaluated the effect of altering presentation of vocabulary definitions (traditional dictionary definition with or without a sample sentence and definitions that were specifically written to be easier to understand) on the learning of novel vocabulary words. In general, regardless of the type of definition given, both the 4th and 6th grade students scored poorly on the task of assessing whether vocabulary usage was consistent with the definition in sentence fragments. However, small but significant gains were found when students were given sample sentences along with the definitions.
- Wu and Solman (1993) investigated the effects of extrapictorial prompts on the learning of words by kindergartners. They found that the best learning occurred equally in two circumstances: in the absence of the pictorial prompts where words were presented alone, and in a feedback cueing condition.
- Eldredge (1990) devised a group-assisted reading method for 3rd grade students. The vocabulary gains for students reading in dyads were greater than for the comparison group of unassisted students who did independent reading.
- Malone and McLaughlin (1997) compared reciprocal peer tutoring with a traditional vocabulary program. The 7th and 8th grade students in the reciprocal peer-tutoring group had significantly higher scores on weekly vocabulary quizzes.

Context Method

The research dealing with contextual approaches to vocabulary acquisition yielded some interesting findings on the role of context and definitional approaches. In accordance with the research findings on rich, extended instruction and multiple exposures to words, one emerging trend was the possibility that the mix of definitional and contextual approaches worked better than either method used alone. Two studies reflect this finding. Kolich (1991) provided computer-assisted practice for 11th grade students; those receiving mixed instruction (context optional word choices and definitional) scored highest. Similarly, Stahl (1983) found those 5th grade students receiving a mixed treatment (definitional and contextual) outscored both students receiving the definitional alone and the students in the control conditions.

However, some studies found specific gains using a single approach. Margosein, Pascarella, and Pflaum (1982) worked with junior high school students and found significant effects for semantic mapping over context-rich or target-word treatment; their work suggests that students should focus on word with similarities to other known words. Gipe and Arnold (1979) compared several vocabulary methods for 3rd and 5th grade students: instruction from context, association, dictionary, and category. They found the highest gains for the context method.

Several studies demonstrated that direct instruction in learning word meanings was helpful for vocabulary acquisition.

- Tomesen and Aarnoutse (1998) included vocabulary instruction for 4th grade students in a program of reciprocal teaching. Students were given direct instruction in deriving word meanings from context. This was found to be more helpful for poor than for average readers, but there was no transfer to general reading comprehension
- White, Graves, and Slater (1990) explored the need for assisting minority or disadvantaged children in grades 1 through 4 and found that direct instruction

in meaning and decoding may help them to an extent.

- Dole, Sloan, and Trathen (1995) worked with 10th grade students on an "alternative" vocabulary treatment condition: teaching students how to select relevant words, learn them on a deep level, and discuss them. These students outscored students taught with the traditional conditions in which students did not learn to this criterion or discuss the words in context.
- Rinaldi, Sells, and McLaughlin (1997) worked with 10.8- to 11.5-year-old students and 3rd graders with reading difficulties to examine effectiveness of a drill and practice intervention on sight word acquisition. During the intervention, all the students more than doubled their correct rates in oral reading and reduced their numbers of errors.
- Dana and Rodriguez (1992) studied the effects of the TOAST (test, organize, anchor, say, test) method of vocabulary learning as compared to various student-selected methods of vocabulary instruction among 6th grade students. They found that students using the TOAST method scored higher than those using student-selected methods on measures of both immediate and delayed retention of words.
- Stump, Lovitt, Fister, Kemp, Moore, and Schroeder (1992) assessed the effects of a precision teaching intervention for general and special education. Assessments of timed vocabulary quizzes supported the finding that the majority of students in the study scored higher on measures of accuracy and fluency.

Results and Discussion

Measurement of Vocabulary

What is available on the issue of measuring vocabulary, despite the noted research gap, is some implicit evidence, which the Panel provided in a breakdown of the types of measures that have been used by researchers studying vocabulary. To obtain this information the Panel tallied, for each study, whether the vocabulary assessment instrument was standardized or experimenter-generated. In some of the studies, vocabulary was assessed with a pretest as well as a posttest. It was possible to determine what types of assessments (standardized of experimenter-generated) were used in 37 of the studies as dependent variables. Figure 2 on the next page shows the distribution of studies in the database as a function of the type of assessment used.

There were six studies that used standardized assessments as the only dependent variable. One of these studies used two measures. There was almost no overlap in the type of standardized measures used, with six different instruments represented.

One other feature in the data was that of the 50 studies coded, 32 administered pretests. Of these 32 studies, 17 used standardized tests. There were 11 different instruments represented in the total.

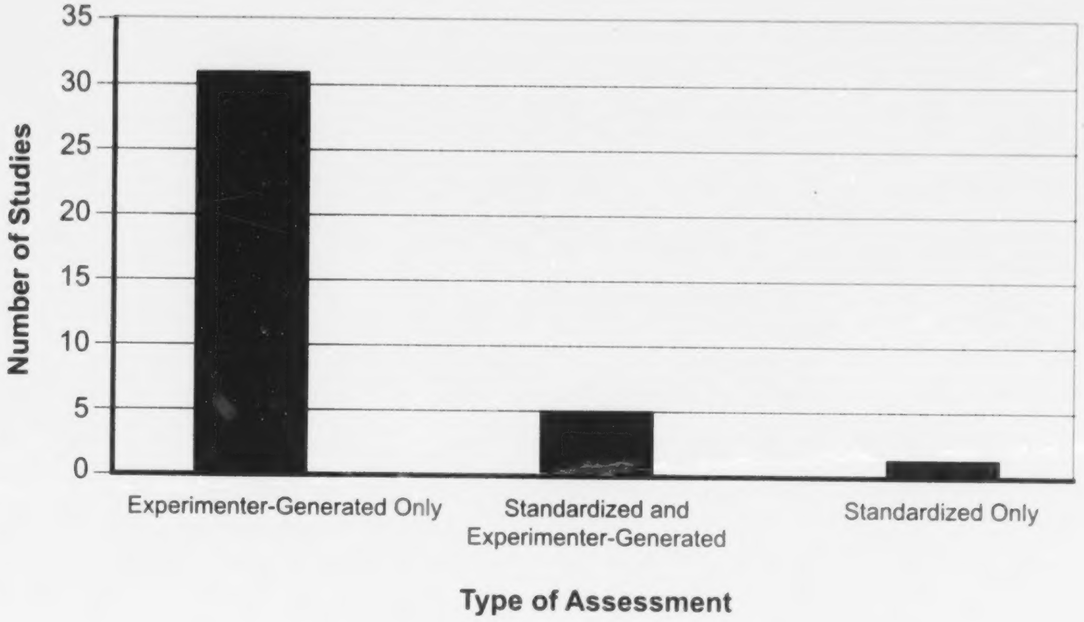
These analyses seem to suggest two implications that might be drawn for practice. First, the standardized tests did not seem to be sufficiently sensitive to vocabulary changes to be used as dependent measures. For practice, this would suggest that assessing vocabulary growth would be best done with teacher-generated instruments as at least one component of evaluation. It also suggests that there may be a need for the development of standardized measures that are much more sensitive to the nuances and complexities involved in vocabulary acquisition. A further implication is that standardized instruments appear to be useful for general screening pretests. Again, the implication for practice might be that standardized tests could be used to identify students who need vocabulary instruction. However, a note of caution is critical here. These implications are tentative and need to be researched before being implemented.

Despite the relatively small body of data available, the collective body of research clearly indicates that vocabulary increases with instruction of many different sorts.

Direct and Indirect Instruction

It is clear that vocabulary should be taught both directly and indirectly. Vocabulary instruction should be incorporated into reading instruction. There is a need for direct instruction of vocabulary items that are required for a specific text to be read as part of the lesson. Direct instruction was found to be highly effective for vocabulary learning (Tomeson & Aarnoutse, 1998; White, Graves, & Slater, 1990; Dole,

Figure 2. Distribution of Assessment Types Included in Database (N = 37 studies)



Sloan, & Trathen, 1995; Rinalid, Sells, & McLaughlin, 1997). In addition, the more connections that can be made to a specific word, the better it seems to be learned. For example, there is empirical evidence indicating that making connections with other reading material or oral language in other contexts seems to have large effects.

Pre-instruction of vocabulary in reading lessons can have significant effects on learning outcomes (Brett, Rothlein, & Hurley, 1996; Wixson, 1986; Carney, Anderson, Blackburn, et al., 1984). At least, it guarantees that there will be fewer unfamiliar concepts in the material to be read. It also helps in making the translation of print to speech meaningful by trying to guarantee that the vocabulary items are in the oral language of the reader. Because almost all early reading is based on oral language, this is a critically important implication.

Repetition and Multiple Exposures

It also seems clear from the Panel's data set that having students encounter vocabulary words often and in various ways can have a significant effect (Senechal, 1997; Leung, 1992; Daniels, 1994, 1996; Dole, Sloan, & Trathen, 1995). Although not a surprising finding, this does have direct implications for instruction. Students should not only repeat vocabulary items in learning; they should be given items that will be likely to appear in many other contexts.

Context

In much the same way that multiple exposures are important, the context in which a word is learned is critical (McKeown, Beck, Omanson, and Pople, 1985; Kameenui, Carnine, & Freschi, 1982; Dole, Sloan, & Trathen, 1995). Vocabulary words should be words that the learner will find useful in many contexts. To that end, a large portion of vocabulary items should be derived from content learning materials. This would serve at least two functions: first, it would assist the learner in dealing with the specific reading matter in content area materials; second, it would provide the learner with vocabulary that would be encountered sufficiently often to make the learning effort worthwhile.

Task Restructuring

Direct vocabulary instruction often assumes that the learner is fully aware of what the task is and how to complete it. However, restructuring tasks can ensure this. Some empirical research has demonstrated the efficacy of being certain that students fully understand the task and the components of vocabulary learning, rather than creating a focus only on the words to be learned (Schwartz & Raphael, 1985). Restructuring the task, such as group learning or revising learning materials, can also lead to increased vocabulary learning (Kameenui, Carnine, & Freschi, 1982; Gordon, Schumm, Coffland, and Doucette, 1992; Wu & Solman, 1993; Eldredge, 1990; Malone & McLaughlin, 1997). This seems to be most effective for low-achieving or at-risk students.

Active Engagement

The few studies that addressed active engagement in learning all reported results consistent with conventional wisdom about learning: Active learning is best. When students were engaged in the tasks in which they were learning vocabulary, they had larger gains (Dickinson & Smith, 1994; Senechal, 1997; Drevno et al., 1994; Daniels, 1994, 1996). This suggests that vocabulary learning tasks that advance other knowledge would be more effective.

Computer Technology

While the use of computer technology in reading is still in its infancy, the few studies reported in the literature suggest that this may be a powerful way of increasing vocabulary (Reinking & Rickman, 1990; Heise et al., 1991; Davidson, Elcock, & Noyes, 1996; Heller, Sturner, Funk & Feezor, 1993). Two possibilities arise here. The first is that the computer might be used as an adjunct to direct vocabulary instruction. In this way, students could obtain more practice in learning vocabulary. A second possibility is that computer technology could bring to bear many different media. This is one way of adding a number of different modalities to the teaching of vocabulary and, consequently, helping ensure more effective vocabulary learning. The availability of online access to vocabulary definitions combines both of these possibilities.

Implicit Learning

It is both a theoretical and an empirical fact that not all vocabulary can or must be learned through formal instruction and that vocabulary words can also be learned through incidental and indirect ways (Robbins & Ehri, 1994; Leung, 1992; Senechal & Cornell, 1993; Nicholson & Whyte, 1992; Stewart et al., 1997). Estimates of vocabulary size seem to suggest that there would never be sufficient classroom time to instruct students to the level of their acquired vocabulary. This implies that much of a student's vocabulary will have to be learned in the course of doing things other than explicit vocabulary learning. Students may well pick up vocabulary in contexts different from the formal learning of a classroom reading group. It may even be that the vocabulary acquired in this way is more memorable, given the role of motivation in its acquisition because the vocabulary acquired in this way may be far more useful. Repetition, richness of context, and motivation may also add to the efficacy of incidental learning.

Assessment and Evaluation of Vocabulary

Although there is no research in the NRP database that bears *directly* on the issue of how vocabulary is assessed, the Panel believes that the way vocabulary is measured can have differential effects on instruction. The Panel bases this belief on several things. First, the plethora of ways in which vocabulary was measured and evaluated in the studies in our database clearly indicates that there is no single standard. Consequently, the Panel suggests that using more than a single measure of vocabulary is critical for sound evaluation. Second, each way of measuring vocabulary produces different results. Furthermore, the category of vocabulary being measured varies. Receptive vocabulary is clearly different from productive vocabulary, and sight vocabulary is yet another concept. Finally, the fact that the Panel found most of the researchers using their own instruments to evaluate vocabulary suggests the need for this to be adopted in pedagogical practice. That is, the more closely the assessment matches the instructional context, the more appropriate the conclusions about the instruction will be.

Standardized tests provide a global measure of vocabulary and may be used to provide a baseline. Few researchers depended on standardized instruments to assess the efficacy of the instruction they studied. The implication for practice is the same: instruments that match the instruction will provide better information about the specific learning of the students related directly to that instruction. The implications for the use of standardized instruments need to be viewed as tentative until the findings can be confirmed by instructional research.

Single vs. Multiple Methods of Instruction

The Panel is reluctant to suggest a single method of learning vocabulary because there were rarely more than a few studies on each individual method. The categories represented in the earlier discussion and the summary of specific methods in Table 1 (Appendix A) reinforce this point. A comprehensive analysis of the collective research studies suggests that a variety of direct and indirect methods of vocabulary instruction can be effective. Effective instructional methods emphasized multimedia aspects of learning, richness of context in which words are to be learned, active student participation, and the number of exposures to words that learners will receive.

Moreover, the age and ability effects discussed above suggest that different methods may be differentially effective. In light of this, dependence on a single method would be a risky course of action.

Implications for Reading Instruction

Based on these trends in the data, the Panel offers the following implications for practice:

1. Vocabulary should be taught both directly and indirectly.
2. Repetition and multiple exposures to vocabulary items are important.
3. Learning in rich contexts is valuable for vocabulary learning.
4. Vocabulary tasks should be restructured when necessary.
5. Vocabulary learning should entail active engagement in learning tasks.
6. Computer technology can be used to help teach vocabulary.
7. Vocabulary can be acquired through incidental learning.
8. How vocabulary is assessed and evaluated can have differential effects on instruction.
9. Dependence on a single vocabulary instruction method will not result in optimal learning.

Directions for Further Research

The following questions do not seem to have clear answers in the research reviewed for this report. They are questions at a relatively high level of generality and are not, in the present form, researchable. That is, they need to be translated into the appropriate variables, operations, and data collection techniques before research can be conducted.

The need in vocabulary instruction research is great. Our knowledge of vocabulary acquisition exceeds our knowledge of pedagogy. That is, the Panel knows a great deal about the ways in which vocabulary increases under highly controlled conditions, but the Panel knows much less about the ways in which such growth can be fostered in instructional contexts. There is a great need for the conduct of research on these topics in authentic school contexts, with real teachers, under real conditions.

1. What are the best ways to evaluate vocabulary size, use, acquisition, and retention? What is the role of standardized tests, what other measures should be used, and under what circumstances?
2. Given the preliminary findings that age and ability levels can affect the efficacy of various vocabulary instruction methods (Tomesen & Aarnoutse, 1998; Robbins & Ehri, 1994; Nicholson & Whyte, 1992; McGivern & Levin, 1983), what are the specific vocabulary instruction needs of students at different grade and ability levels?
3. What are the more general effects of vocabulary instruction across the grades?
4. Empirical support has been found for the facilitation of vocabulary learning with computers as ancillary aids and replacements of other technologies (Reinking & Rickman, 1990; Heise, 1991;

Davidson, Elcock, & Noyes, 1996). What is the optimal use of computer and other technologies in vocabulary instruction? What is the precise role of multimedia learning in vocabulary acquisition?

5. What is the precise role of multimedia learning in vocabulary instruction across the grades?
6. How should vocabulary be integrated into comprehension instruction for optimal benefit to the student?
7. What are the optimal combinations of the various methods of vocabulary instruction, including direct and indirect instruction, and of different methods within these categories?
8. What sort of professional development is needed for teachers to become proficient in vocabulary instruction?

References

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74(4), 506-521.

Blachowicz, C., & Fisher, P. (in press). Vocabulary Instruction. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr, (Eds.). Handbook of reading research (Vol. 3). Mahwah, NJ: Lawrence Erlbaum Associates.

Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. Elementary School Journal, 96(4), 415-422.

Carney, J. J., Anderson, D., Blackburn, C., & Blessing, D. (1984). Preteaching vocabulary and the comprehension of social studies materials by elementary school children. Social Education, 48(3), 195-196.

Dana, C., & Rodriguez, M. (1992). TOAST: A system to study vocabulary. Reading Research and Instruction, 31(4), 78-84.

Daniels, M. (1994). The effect of sign language on hearing children's language development. Communication Education, 43(4), 291-298.

Daniels, M. (1996). Bilingual, bimodal education for hearing kindergarten students. Sign Language Studies, 90, 25-37.

Davidson, J., Elcock, J., & Noyes, P. (1996). A preliminary study of the effect of computer-assisted practice on reading attainment. Journal of Research in Reading, 19(2), 102-110.

Davis, F. B. (1942). Two new measures of reading ability. Journal of Educational Psychology, 33, 365-372.

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. Reading Research Quarterly, 29, 104-122.

Dole, J. A., Sloan, C., & Trathen, W. (1995). Teaching vocabulary within the context of literature. Journal of Reading, 38(6), 452-460.

Drevno, G. E., Kimball, J.W., Possi, M. K., Heward, W. L., Gardner, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and generalization of science vocabulary by elementary students: A systematic replication. Journal of Applied Behavior Analysis, 27(1), 179-180.

Durkin, D. (1993). Teaching them to read. (6th ed.) Boston, MA: Allyn and Bacon.

Eldredge, J. L. (1990). Increasing the performance of poor readers in the third grade with a group-assisted strategy. Journal of Educational Research, 84(2), 69-77.

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. Review of Educational Research, 68(4), 450-469.

Gipe, J. P., & Arnold, R. D. (1979). Teaching vocabulary through familiar associations and contexts. Journal of Reading Behavior, 11(3), 281-285.

Gordon, J., Schumm, J. S., Coffland, C., & Doucette, M. (1992). Effects of inconsiderate vs. considerate text on elementary students' vocabulary learning. Reading Psychology, 13(2), 157-169.

Heise, B. L., Papalewis, R., & Tanner, D. E. (1991). Building base vocabulary with computer-assisted instruction. Teacher Education Quarterly, 18(1), 55-63.

Heller, J. H., Sturner, R. A., Funk, S. G., & Feezor, M. D. (1993). The effect of input mode on vocabulary identification performance at low intensity. Journal of Educational Computing Research, 9(4), 509-518.

Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. Reading Research Quarterly, 24(2), 215-235.

Kameenui, E., Carnine, D., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. Reading Research Quarterly, 17(3), 367-388.

Klesius, J. P., & Searls, E. F. (1990). A meta-analysis of recent research in meaning vocabulary instruction. Journal of Research and Development in Education, 23(4), 226.

Kolich, E. M. (1991). Effects of computer-assisted vocabulary training on word knowledge. Journal of Educational Research, 84(3), 177-182.

Koury, K. A. (1996). The impact of preteaching science content vocabulary using integrated media for knowledge acquisition in a collaborative classroom. Journal of Computing in Childhood Education, 7(3-4), 179-197.

Leung, C. B. (1992). Effects of word-related variables on vocabulary growth repeated read-aloud events. In C. K. Kinzer & D. J. Leu (Eds.), Literacy research, theory, and practice: Views from many perspectives: Forty-first Yearbook of the National Reading Conference (pp. 491-498). Chicago, IL: The National Reading Conference.

Levin, J. R., Levin, M. E., Glasman, L. D., & Nordwall, M. B. (1992). Mnemonic vocabulary instruction: Additional effectiveness evidence. Contemporary Educational Psychology, 17(2), 156-174.

Levin, J., Johnson, D., Pittelman, S., Levin, K., Shriberg, L., Toms-Bronowski, S., & Hayes, B. (1984). A comparison of semantic- and mnemonic-based vocabulary-learning strategies. Reading Psychology, 5(1-2), 1-15.

Levin, J., McCormick, C., Miller, G., & Berry, J. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies for children. American Educational Research Journal, 19(1), 121-136.

Malone, R. A., & McLaughlin, T. F. (1997). The effects of reciprocal peer tutoring with a group contingency on quiz performance in vocabulary with seventh- and eighth-grade students. Behavioral Interventions, 12(1), 27-40.

Margosein, C. M., Pascarella, E. T., & Pflaum, S. W. (1982). The effects of instruction using semantic mapping on vocabulary and comprehension. Journal of Early Adolescence, 2(2), 185-194.

McFalls, E. L., Schwanenflugel, P. J., & Stahl, S. A. (1996). Influence of word meaning on the acquisition of a reading vocabulary in second-grade children. Reading and Writing, 8(3), 235-250.

McGivern, J. E., & Levin, J. R. (1983). The keyword method and children's vocabulary learning: An interaction with vocabulary knowledge. Contemporary Educational Psychology, 8(1), 46-54.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15(1), 3-18.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. Reading Research Quarterly, 20(5), 522-535.

Medo, M. A., & Ryder, R. J. (1993). The effects of vocabulary instruction on readers' ability to make causal connections. Reading Research and Instruction, 33(2), 119-134.

Meyerson, J. J., Ford, M. S., & Jones, W. P. (1991). Science vocabulary knowledge of third and fifth grade students. Science Education, 75(4), 419-428.

Nagy, W., & Scott, J. (in press). Vocabulary processes. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr, (Eds.), Handbook of reading research (Vol. 3). Mahwah, NJ: Lawrence Erlbaum Associates.

Nicholson, T., & Whyte, B. (1992). Matthew effects in learning new words while listening to stories. In C. K. Kinzer & D. J. Leu (Eds.), Literacy research, theory, and practice: Views from many perspectives: Forty-first Yearbook of the National Reading Conference (pp. 499-503). Chicago, IL: The National Reading Conference.

Reinking, D., & Rickman, S. S. (1990). The effects of computer-mediated texts on the vocabulary learning and comprehension of intermediate-grade readers. Journal of Reading Behavior, 22(4), 395-411.

Rinaldi, L., Sells, D., & McLaughlin, T. F. (1997). The effects of reading racetracks on the sight word acquisition and fluency of elementary students. Journal of Behavioral Education, 7(2), 219-233.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. Journal of Educational Psychology, 86(1), 54-64.

Ryder, R. J., & Graves, M. F. (1994). Vocabulary instruction presented prior to reading in two basal readers. Elementary School Journal, 95(2), 139-153.

Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997) Partial word knowledge and vocabulary growth during reading comprehension. Journal of Literacy Research, 20, 531-553.

Schwartz, R. M., & Raphael, T. E. (1985). Instruction in the concept of definition as a basis for vocabulary acquisition. In J. A. Niles & R.V. Lalik (Eds.), Issues in literacy: A research perspective: Thirty-fourth Yearbook of the National Reading Conference (pp. 116-124). Rochester, NY: The National Reading Conference.

Scott, J. A., & Ehri, L. C. (1990). Sight word reading in prereaders: Use of logographic vs. alphabetic access routes. Journal of Reading Behavior, 22(2), 149-166.

Scott, J., & Nagy, W. (1997). Understanding the definitions of unfamiliar verbs. Reading Research Quarterly, 32, 184-200.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. Journal of Child Language, 24(1), 123-138.

Senechal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. Reading Research Quarterly, 28(4), 360-374.

Stahl, S. (1983). Differential word knowledge and reading comprehension. Journal of Reading Behavior, 15(4), 33-50.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. Review of Educational Research, 56(1), 72-110.

Stahl, S. A., Richek, M. A., & Vandevier, R. J. (1991). Learning meaning vocabulary through listening: A sixth-grade replication. In J. Zutell & S. McCormick (Eds.), Learner factors/teacher factors: Issues in literacy research and instruction: Fortieth Yearbook of the National Reading Conference (pp. 185-192) Chicago, IL: The National Reading Conference.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. Journal of Educational Psychology, 83(1), 8-16.

Stewart, S. R., Gonzalez, L. S., & Page, J. L. (1997). Incidental learning of sight words during articulation training. Language, Speech, and Hearing Services in Schools, 28(2), 115-125.

Stump, C. S., Lovitt, T. C., Fister, S., Kemp, K., Moore, R., & Shroeder, B. (1992). Vocabulary intervention for secondary-level youth. Learning Disability Quarterly, 15(3): 207-222.

Tomesen, M., & Aarnoutse, C. (1998). Effects of an instructional programme for deriving word meanings. Educational Studies, 24(1), 107-128.

Whipple, G. (Ed.). (1925). The Twenty-fourth Yearbook of the National Society for the Study of Education: Report of the National Committee on Reading. Bloomington, IL: Public School Publishing Company.

White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. Journal of Educational Psychology, 82(2), 281-290.

Wixson, K. K. (1986). Vocabulary instruction and children's comprehension of basal stories. Reading Research Quarterly, 21(3), 317-329.

Wu, H.-M., & Solman, R. T. (1993). Effective use of pictures as extra stimulus prompts. British Journal of Educational Psychology, 63(1), 144-160.

Appendix A

Vocabulary Instruction Methods

TABLE 1: A SUMMARY OF VOCABULARY INSTRUCTION METHODS

Vocabulary Method:	Description:	Representative Studies:
Keyword Method	Students are instructed to learn the meanings of new words by learning a keyword "word clue" for each vocabulary word. The keywords are usually words acoustically similar to a salient part of the vocabulary word. Sometimes, relational illustrations are shown to students, or students are asked to generate their own images linking the two words.	Levin, Levin, Glassman, & Nordwall, 1992; McCarville, 1993; Levin, Levin, Glassman, & Nordwall, 1992; Levin, Levin, Cotton, Bartholomew, Hasty, Hughes, & Townsend, 1990; Pressley, Levin, Kuiper, Bryant, & Mitchener, 1982; Atkinson, 1975.
Semantic Mapping	Students are taught the meanings of new vocabulary words by categorizing them into familiar topics with other known words. New words are learned by identifying similarities and differences with related, known words. Target words are often introduced in categories, and semantic maps are developed for each set of items.	Levin, Levin, Glassman, & Nordwall, 1992; McCarville, 1993; Levin, Levin, Glassman & Nordwall, 1992; Levin, Levin, Cotton, Bartholomew, Hasty, Hughes, & Townsend, 1990; Pressley, Levin, Kuiper, Bryant, & Mitchener, 1982; Atkinson, 1975.
Contextual Analysis	Students use context clues embedded in paragraphs to help them learn meanings of the target vocabulary words. Usually, the words and definitions are then reviewed.	Buikema & Graves, 1993; Friedl&, 1992; Gifford, 1993.
Sign Language	Enrichment of curriculum with sign language for pre-kindergarten hearing children to improve their receptive English vocabulary.	Daniels, 1994, 1998.
Wide Reading	Listening/reading stories (with or without pre-explanation of target words). Some salient variables to consider include the number of exposures to the words, frequency of book readings, nature of instruction (using questioning, etc.), word redundancy, and time between readings.	Anderson & Nagy, 1992; Riddell, 1988; Elley, 1988; Krashen, 1989.
Deriving Word Meanings	Students are taught strategies for deriving meaning of an unfamiliar word. One example of a strategy is the SCANR method (substitute a word for unknown word; check the context for clues; ask if substitution fits context clues; need a new idea?; revise idea to fit context)	Tomeson, 1998; Jenkins, Matlock, & Slocum, 1989.
Elaborate/Rich Instruction	Students learn to identify the relationship between words, respond to words both affectively and cognitively, and apply words to various contexts. Promotes a student's use of words outside of vocabulary class and elicits prior knowledge.	McKeown, Beck, Omanson, & Pople, 1985; Stanley & Ginther, 1991; Stahl, 1983.
Roots/Affix Analysis	Students use word origin clues and learn the meanings of common roots, prefixes, and affixes to determine vocabulary definitions.	Irwin, 1991, Ryder & Graves, 1994; Levin, Carney, & Pressley, 1988.

BEST COPY AVAILABLE

TABLE 1: A SUMMARY OF VOCABULARY INSTRUCTION METHODS (CONTINUED)

Vocabulary Method:	Description:	Representative Studies:
Dictionary/Glossary	Students are given dictionaries or glossaries to find the definitions of unknown words. Variations of this method include giving students passages to read along with a dictionary or glossary to find the definitions of unknown words, writing new sentences with the words, and completing worksheets and crossword puzzles.	Knight, 1994; Wixson, 1986; Gipe & Arnold, 1979.
Frayer Model	Frayer model (and Grave's modification). A method to teach specific new words using a seven-step model. Basic tenets of the Frayer model include: give word/name and its relevant attributes, eliminate irrelevant attributes, give examples, give nonexamples, and list subordinate, superordinate, and coordinate terms.	Frayer, Frederick, & Klausmeier, 1969; Graves, 1984, 1985; Ryder & Graves, 1994.
Task Clarification	With the premise that students have only a vague notion of what constitutes a definition, vocabulary instruction is designed to clarify the student's knowledge of the task. Students are instructed on ways to gather information from relevant sources to uncover the components of a definition.	Guzzetti, Snyder, Glass, & Gamas, 1993; Haggard, 1982, 1985; Fisher, Blachowicz, & Smith, 1991; Fisher & Danielsen, 1998; Palinscar & Brown, 1984.
Computer/Multimedia Instruction	Various methods incorporate computer and multimedia technology to aid in the instruction of vocabulary words. Examples include CD-ROM, talking software, Hypertext dictionary support, speech prompts, adaptive software, visual representations, and multisensory input.	Terrell & Daniloff, 1996; Reinking & Rickman, 1990.
Text Revision	Students are given revised versions of text passages. Variations include substituting easy for difficult vocabulary words, adding redundant information to facilitate word learning and comprehension, and writing vocabulary words with context information to constrain vocabulary word learning.	Britton, Woodward, & Binkley, 1993; Meyer, 1975.
Interactive Vocabulary Techniques	Various techniques that allow students to get actively involved in word learning. Examples include students acting out word meanings, self-selection of vocabulary words to learn, and allowing students to compare strategies and methods.	Duffelmeyer, 1980; Rekrut, 1993; Pressley & Levin, 1988.
Passage Integration Training	Teachers stop and prompt the students to generate the meanings of the difficult vocabulary words immediately after they encounter them during the passage reading.	Kameenui, Carnine, & Freschi, 1982.
Concept Method	Assists students in learning words as concepts rather than as dictionary definitions. Based on a concept-attainment model, this method relies more heavily on discussion than on independent activities. Students study examples and nonexamples to identify the critical attributes of each word or concept.	Frayer et al., 1969; Klausmeier, 1976,1979; Merrill & Tennyson, 1977.
Pre-Instruction of Vocabulary Words	Students are taught or exposed to the definitions of relevant vocabulary words before reading them in context. In addition to assessing effects on vocabulary acquisition, this is often researched as a way to enhance reading comprehension.	Koury, 1996; Ryder & Graves, 1994; Wixson, 1986.

TABLE 1: A SUMMARY OF VOCABULARY INSTRUCTION METHODS (CONTINUED)

Vocabulary Method:	Description:	Representative Studies:
Association Methods	Pairs unknown word with familiar synonym. Students must memorize the pairings to rewrite the original pairs.	Gipe & Arnold 1979; McKeown, Beck, Omanson, & Pople, 1985.
TOAST Program	Students are taught a method of vocabulary instruction by the acronym of TOAST that prompts students to: test, organize words, anchor words, and test target words.	Dana & Rodriguez, 1992.
Basic Mnemonic Techniques	Traditional memory techniques, including vocabulary drills, flash cards, vocabulary games, notebooks, repetitions, and recall tests. An example program is the Reading Racetrack, which uses error correction, timing, and drill and practice procedures to help build sight word acquisition and reading fluency.	McLaughlin, 1997; Rinaldi, Sells, & McLaughlin, 1997.
Decoding Instruction	To enhance reading fluidity with the intention of facilitating vocabulary comprehension, instruction is given in methods such as phonological training, phonemic awareness, or the whole-word approach.	Eldredge, Quinn, & Butterfield, 1990.

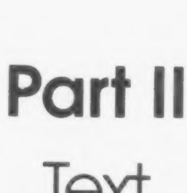
BEST COPY AVAILABLE

CHAPTER 4

Comprehension

Part II

Text Comprehension Instruction



COMPREHENSION II

Text Comprehension Instruction

Introduction

An examination of the scientific basis for instruction of text comprehension was undertaken by members of the NRP. The Panel decided to focus on instruction of vocabulary, on instruction of comprehension of text, and on the preparation of teachers to teach comprehension of text. This report presents a review of the scientific evidence on the instruction of comprehension of text in normal readers.

Comprehension has come to be viewed as "the essence of reading" (Durkin, 1993). Although comprehension of text is now regarded as essential to reading and learning, comprehension as a process began to receive scientific attention only in the past 30 years. Beginning in the 1970s, researchers such as Markman (1977, 1981) began to study the awareness that readers had of their comprehension processes during reading. The questions were whether readers knew that they did not understand what they were reading in a text and what they did if they recognized that they had an understanding failure. The initial, surprising finding by Markman was that both young and mature readers failed to detect logical and semantic inconsistencies in the text. This discovery of comprehension failure led to the identification and teaching of strategies that readers could learn to enhance their comprehension (see below).

An important development in theories about reading comprehension occurred in the 1970s. Reading comprehension was seen not as a passive, receptive process but as an active one that engaged the reader. Reading came to be seen as intentional thinking during which meaning is constructed through interactions between text and reader (Durkin, 1993). According to this view, meaning resides in the intentional, problem-solving, thinking processes of the reader that occur during an interchange with a text. The content of meaning is influenced by the text and by the reader's prior knowledge that is brought to bear on it (Anderson & Pearson, 1984). Reading comprehension was seen as the construction of the meaning of a written text through a reciprocal interchange of ideas between the reader and the message in a particular text (see, for example, Harris & Hodges, 1995, definition #2, p. 39). The important theoretical idea here was that readers construct meaning representations of the text as they read and that these representations were essential to memory and use of what was read and understood. This view was furthered by the publication of important papers on dynamic models of the comprehension processes such as that by Kintsch and van Dijk (1978). Here, readers were assumed to construct mental representations of what they read. These representations were stored in memory and contained the semantic interpretations of the text made by the reader during reading. The memory representations provided the basis for subsequent use of what was read and understood.

The bulk of instruction of text comprehension research during the past 3 decades has been guided by this cognitive conceptualization of reading. In the cognitive research of the reading process, reading is purposeful and active (Pressley & Afflerbach, 1995). According to this view, a reader reads a text to understand what is read, to construct memory representations of what is understood, and to put this understanding to use. A reader can read a text to learn, to find out information, or to be entertained. These various purposes of understanding require that the reader use knowledge of the world, including language and print. This knowledge enables the reader to make meaning of the text, to form memory representations of these meanings, and to use them to communicate with others information about what was read.

Although instruction on text comprehension has been a major research topic for more than 20 years, the explicit teaching of text comprehension before the 1970s was done largely in content areas and not in the context of formal reading instruction (Durkin, 1979). The idea behind explicit instruction of text comprehension is that comprehension can be improved by teaching students to use specific cognitive strategies or to reason strategically when they encounter barriers to comprehension when reading. The goal of such training was the achievement of competent and self-regulated reading.

Readers normally acquire strategies for active comprehension informally. Comprehension strategies are specific procedures that guide students to become aware of how well they are comprehending as they attempt to read and write. Explicit or formal instruction on these strategies is believed to lead to improvement in text understanding and information use. Instruction in comprehension strategies is carried out by a classroom teacher who demonstrates, models, or guides the reader on their acquisition and use. When these procedures have been acquired, the reader becomes independent of the teacher. Using them, the reader can effectively interact with the text without assistance. Readers who are not explicitly taught these procedures are unlikely to learn, develop, or use them spontaneously.

The past 30 years of the scientific study of instruction of text comprehension reveal a distinct trend. The initial investigations focused on the training of particular individual strategies such as comprehension monitoring or identifying main ideas. Here the question was whether readers could learn to use an individual strategy. Then, the focus was on whether particular strategies could be learned and whether they could facilitate comprehension. This was an important advance because it validated the teaching of text comprehension strategies. Next, researchers began to study whether the teaching of combinations of different strategies lead to their acquisition and improvement of text comprehension. The success of these "multiple" strategy teaching methods led to study of the preparation of teachers to teach strategies in natural classroom contexts. This historical development from the instruction of individual strategies to the preparation of teachers to implement them in interaction with readers in the classroom is an important contribution of the scientific approach to the study of reading instruction. The Panel's review covers this history of instruction of text comprehension.

Cognitive Strategies for Improving Reading Comprehension

Comprehension strategies are procedures that guide students as they attempt to read and write. For example, a reader may be taught to generate questions about the text as it is read. These questions are of the why, what, how, when, or where variety; and by generating and trying to answer them, the reader processes the text more actively. The value of cognitive strategies in comprehension instruction is, first, their usefulness in the development of instructional procedures, and second, the learning of these procedures by students as an aid in their reading and learning, independent of the teacher.

Instruction of strategies for comprehending during reading is a way for teachers to break though students' passivity and involve them in their own learning (Mier, 1984). Typically, instruction of cognitive strategies employed during reading consists of:

1. The development of an awareness and understanding of the reader's own cognitive processes that are amenable to instruction and learning
2. A teacher guiding the reader or modeling for the reader the actions that the reader can take to enhance the comprehension processes used during reading
3. The reader practicing those strategies with the teacher assisting until the reader achieves a gradual internalization and independent mastery of those processes (Palinscar & Brown, 1984; Paris & Oka, 1986; Pressley et al., 1994).

The general finding is that when readers are given cognitive strategy instruction, they make significant gains on measures of reading comprehension over students trained with conventional instruction procedures (Pressley et al., 1989; Rosenshine & Meister, 1994; Rosenshine, Meister, & Chapman, 1996).

From a historical perspective, instruction in how to comprehend is not new. Benjamin Franklin invented a "weighted characteristics test" used in a current instruction curriculum for readers to apply for making decisions about ideas in texts while reading (Block, 1993). E. L. Thorndike claimed back in 1917 that "reading is reasoning." Despite Thorndike's arguments, however, beginning readers were seldom taught cognitive strategies that could assist them in reading. Durkin's (1979) highly cited observational studies of reading instruction in grade 4 showed that teachers, in fact, spent little time on comprehension instruction. Only 20 minutes of comprehension instruction was observed in 4,469 minutes of reading instruction. This lack was echoed by Duffy, Lanier, and Roehler (1980). They described teachers as spending time in assigning activities, supervising and monitoring students as to being on task, directing recitation sessions as a way of assessing what the students were doing, and providing corrective feedback when the students erred. The teachers did not teach or show the students skills, strategies, or processes that they could use in reading to comprehend what they read and to be successful in learning information in the text.

Research on instruction of comprehension strategies that could help students improve their reading comprehension began in the late 1970s and has thrived since. According to Rosenshine, Meister, and Chapman (1996), the earliest uses of the term "comprehension monitoring" is found in Markman (1978, 1979), Gagne (1977), and Weinstein (1978). Researchers and educators have long been interested in what we think about thinking, in how our knowledge develops, and in how what we know about how our own thought processes affect reading comprehension. The focus on what we know about cognition has led to the development of practical strategies for improving students' comprehension. The cumulative result of nearly 3 decades of research is that "there is ample extant research supporting the efficacy of cognitive strategy training during reading as a means to enhance students' comprehension" (Baumann, 1992, p. 162).

Methodology

Database

In order to conduct a scientific review of the research on comprehension instruction during the past 2 decades, the Panel located studies since 1980 by searching the PsycINFO and ERIC databases electronically. The Panel used the terms comprehension, strategy, and instruction. From this search, the Panel identified 453 studies on comprehension. In addition, the Panel added other studies that were from the 1970s or otherwise not revealed in the search. In this regard, reviews or studies on strategy instruction by Duffy and Roehler (1989); Lysynchuk, Pressley, d'Ailly, Smith, and Cake (1989); Pressley, Johnson, Symons, McGoldrick, and Kurita (1989); Pressley (1998); Rosenshine and Meister (1994); and Rosenshine, Meister, and Chapman (1996) proved to be very helpful. As a result, an additional 28 studies not found initially in the electronic search were added to the Panel's review.

Analysis

In order to be included in the NRP's scientific review of the research literature on instruction of text comprehension, a study had to be:

1. Relevant to instruction of reading or comprehension among normal readers. This criterion, in particular, excluded studies on comprehension instruction in reasoning and mathematics problem solving (Schoenfeld, 1985), physics (Larkin & Reif, 1976), and writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985).
2. Published in a scientific journal. A few exceptions are dissertations and conference proceedings that were reviewed in two meta-analyses by Rosenshine and his colleagues (Rosenshine & Meister, 1994; Rosenshine, Meister, & Chapman, 1996).
3. Have an experiment that involved at least one treatment and an appropriate control group or have one or more quasi-experimental variables with variations that served as comparisons between treatments. The latter was rare.
4. In so far as could be determined, have the participants or classrooms randomly assigned to the treatment and control groups or matched on initial measures of reading comprehension. This criterion was relaxed in a number of studies where random assignment of classrooms was not carried out.

The application of these criteria reduced the number of studies to be reviewed from 481 to 205. The Panel then coded and entered the coded contents of these studies into a database to identify the types of comprehension instruction that were reported as effective. Because the studies numbered 205, the Panel first analyzed the abstracts of the studies, coding the kind of instruction, experimental treatments and controls (independent variables), grade and reading level of readers, instructor (teacher or experimenter), assessments (dependent variables), and kind of text. The Panel then classified and grouped studies based upon the kinds of instruction used. The Panel identified 16 distinct categories of instruction. Table 1, on the following page, summarizes the 16 categories of a total of 203 studies that met the NRP criteria for inclusion as scientific studies on comprehension instruction. It shows the type of instruction used, the number of studies using that kind of instruction, a brief rationale as to why instruction was used, and generally whether and how it was effective.

Each category of studies is summarized in Appendix A. The summaries define and describe the rationale for each kind of instructional strategy, the procedures used, and how the instruction is assessed by the researchers. The Panel then evaluated the category of instruction, based on reported results.

In Appendix B, a table summarizes the 16 categories of instruction, describing the effects claimed by the researchers, the grade levels that were studied, and how the method might be taught in a classroom setting.

In order to draw scientific conclusions about a finding, one needs evidence that an experimental effect is reliable, robust, replicable, and general. Reliability of an effect is decided by differences that statistically favor a treatment. Robustness of an effect is determined by the magnitude of effects over replications. Replication is determined by independent validation of significant treatment effects. Generality is determined by the transfer measures. In this review, experimenter tasks reflect near transfer and standardized tests reflect far transfer. The NRP evaluated how well each strategy met these criteria. The main criteria that the NRP used are reliability, replication, and generality. Robustness was not determined in most cases because effect sizes could not be calculated for almost all of the studies. Effect size data, however, were available from two meta-analyses by Rosenshine and his colleagues (Rosenshine & Meister, 1994; Rosenshine et al., 1996).

Consistency With the Methodology of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. A formal meta-analysis was not possible because even the studies identified in the same instructional category used widely varying sets of methodologies and implementations. Therefore, the Panel found few research studies that met all the NRP criteria; however, to the extent possible, NRP criteria were employed in the analyses. An examination of the quality of the research studies appears in the Discussion section of this report. NRP criteria for Evaluating Existing Reviews of Research were used in the analyses of the two Rosenshine and colleagues meta-analyses.

Results

Of the 16 categories of instruction, 7 appear to have a firm scientific basis for concluding that they improve comprehension in normal readers. The seven individual strategies that appear to be effective and most promising for classroom instruction are (in alphabetical order) comprehension monitoring, cooperative learning, graphic and semantic organizers including story maps, question answering, question generation, and summarization. In addition, many of these strategies have also been effectively used in the category "multiple strategy," where readers and teachers interact over texts.

Mental Imagery and Mnemonic (Keyword) Strategies have reliable effects on improving memory for text. These procedures may be useful when teachers wish to use an alternative way of having the reader try to understand and represent text. These procedures are useful for recall of individual sentences or paragraphs.

Curriculum-Plus-Strategies, Psycholinguistic, and Listening Actively studies were so few that an assessment of the scientific merit of a particular treatment could not be made. The use of instructional procedures that activate prior knowledge was found to be quite varied. The activation of prior knowledge may be obtained through other means such as question elaboration, question generation, or question answering as well as other forms of content area exposure such as teacher lectures, films, and discussion before reading.

Two categories on which there were few studies have, in the view of the NRP, considerable promise for future study. Only four studies were found on the Preparation of Teachers on comprehension instruction strategies. These studies are important because they represent a culmination in the evolution of text comprehension

Table 1
Categories of Comprehension Instruction

TYPE OF INSTRUCTION	# OF STUDIES	WHY INSTRUCT?	HOW EFFECTIVE?
Comprehension Monitoring	22	Readers do not show comprehension strategy awareness.	Readers learn to monitor how well they comprehend.
Cooperative Learning	10	Readers need to learn to work in groups, listen and understand their peers as they read, and help one another use strategies that promote effective reading comprehension.	Readers learn to focus and discuss reading materials. Readers learn reading comprehension strategies and do better on comprehension tests. Teachers provide cognitive structure.
Curriculum	8	Strategies should be integrated into the normal curriculum.	Readers improve reading ability and academic achievement.
Graphic Organizer	11	Readers do not use external organization aids that can benefit their understanding.	Readers improve memory and comprehension for text.
Listening Actively	4	Readers do not listen effectively.	Readers improve memory and comprehension for text.
Mental Imagery	7	Readers do not use imagery.	Readers improve memory and comprehension for text.
Mnemonic	2	Pictorial aids are not usually available; and these, plus keywords, help readers learn and organize information.	Readers improve memory and comprehension for text.
Multiple Strategies	38	Readers need to learn to coordinate several processes in order to construct meaning from texts.	Readers improve reading ability and academic achievement.
Prior Knowledge	14	Readers may not have relevant knowledge during reading.	Readers improve memory and comprehension for text.
Psycholinguistic	1	Reader may lack relevant knowledge about language.	Readers learn to identify antecedents of pronouns.

TABLE 1
CATEGORIES OF COMPREHENSION INSTRUCTION (continued)

TYPE OF INSTRUCTION	# OF STUDIES	WHY INSTRUCT?	HOW EFFECTIVE?
Question Answering	17	Readers do not know how to answer questions, nor do they know how to make inferences.	Readers improve answering questions.
Question Generation	27	Readers do not know how to generate questions or inferences.	Readers learn to generate and answer inferential questions.
Story Structure	17	Poor readers cannot identify structure.	Readers improve memory and identification of story structure.
Summarization	18	Readers do not know how to summarize text.	Readers improve memory and identification of main ideas.
Teacher Preparation	6	Teachers do not ordinarily use effective transactional strategies.	Teachers learn strategies. Readers improve reading comprehension.
Vocabulary-Comprehension Relationship	3	Reading comprehension depends upon word knowledge.	Readers learn word meanings and improve comprehension.

instruction during the past 2 decades. These studies also represent essential investigations because in most of the text comprehension strategy instruction reviewed, strategies were taught by experimenters rather than classroom teachers. It is important to know whether strategies can be learned and used faithfully and effectively by teachers in classroom contexts. These four studies are intensively reviewed as a part of the Comprehension report section on teacher preparation.

Success in instruction on the relation of vocabulary to comprehension has been found in only two studies with 8th graders. This is an important kind of instruction that needs to be investigated on a wider range of grade levels. The Panel would like to know what the relationship is between word learning and comprehension. The review on vocabulary in Comprehension I (Vocabulary Instruction) shows that vocabulary can be successfully taught over a wide range of grades.

Comprehension Monitoring meets criteria of reliability and replication for the *specific* learning of the strategy (100% effectiveness in 14 studies across grades 2 through 6). Although comprehension monitoring is believed to be important as a part of a multiple strategy method, the evidence for it alone having a general effect is less compelling. Reliable effects are reported on only three experimenter tasks (error detection, recall, question answering) with two reported failures on 2nd graders. The number of studies reporting the use of transfer tests is small (four on reliable experimenter effects and five on reliable standardized tests). The method does not seem to generalize for 2nd graders. Nevertheless, it may be a useful addition to a program of instruction that employs flexibility and the teaching of multiple comprehension strategies.

Cooperative Learning showed 10 studies that reported reliable effects of instruction on grade levels 3 through 6 on experimenter tasks. Only three studies used standardized tests. Thus, cooperative learning produces reliable and replicable near transfer. The evidence for generalization is based on a small number of studies. Having peers instruct or interact over the use of reading strategies leads to an increase in the learning of the strategies, promotes intellectual discussion, and increases reading comprehension. This procedure saves on teacher time and gives the students more control over their learning and social interaction with peers.

Graphic Organizers were used in 11 studies on texts used in Social Studies and Science. The most frequent grade levels were 4 to 6. Children who can learn and benefit from this instruction have to have skill in writing and reading. The empirical evidence indicates reliable and replicable effects on near transfer tasks of memory for reading content (six of seven studies). The main effect of graphic organizers appears to be on the improvement of the reader's memory for the content that has been read. General effects are reported in four studies on achievement gains in content areas. Although the number is small, success in increasing achievement in a context subject is promising. Only two studies report the use of standardized tests so that evidence is limited in replication on this kind of general transfer. Teaching students to organize the ideas that they are reading about in a systematic, visual graph benefits the ability of the students to remember what they read and may transfer, in general, to better comprehension and achievement in Social Studies and Science content areas. The success here suggests that the instruction of comprehension could be carried out in content area teaching.

Question Answering was investigated in 17 studies, mainly in grades 3 through 5. The evidence is primarily that the effects are specific to increased success on experimenter tests of question answering. There are no reports of standardized or other general tests. This procedure may be best used as a part of multiple strategy packages where the teacher uses questions to guide and monitor readers' comprehension.

Question Generation. The strongest scientific evidence was found for the effectiveness of asking readers to generate questions during reading. There were 27 studies on this treatment that was used on readers in grades 3 through 9 (mode = 6). The main support comes from the large number of studies that assessed effectiveness by both experimenter and standardized tests as well as a meta-analysis by Rosenshine, Meister, and Chapman (1996). In the latter analysis, the respective effect sizes for multiple choice (n = 6), short-answer (n = 14), and summary (n = 3) measures were 0.95, 0.85, and 0.85, respectively. On standardized tests, the median effect size for 13 studies that used standardized comprehension tests was 0.36. Although there is a positive effect size for standardized tests, only 3 of 13 effects were statistically significant, casting doubt on the generality of this single strategy instruction. In contrast, experimenter tests fared better because 16 of 19 were statistically significant. Thus, there was stronger evidence for near transfer than for generalized effects. There is mixed evidence that general reading comprehension is improved on standardized, comprehension tests. Question generation may also be best used as a part of a multiple strategy instruction program.

Story Structure is a procedure used extensively in reading comprehension of narrative texts. There are 17 studies over grades 3 through 6, about one half of which were focused on poor readers. The success in the treatment is more frequent with poor or below-average readers; good readers do not seem to need this kind of instruction. The treatment successfully transfers to question answering and recall. Only a few (two of three) studies report transfer to standardized comprehension tests. The instruction of the content and organization of stories thus improves comprehension of stories as measured by the ability of the reader to answer questions and recall what was read. This improvement is more marked for less able readers. More able readers may already know what a story is about and therefore do not benefit as much from the training. However, this kind of instruction may aid both kinds of readers in terms of writing as well as reading literary texts. Because stories are used extensively in elementary school, instruction on how to understand a story is warranted by the data, especially for less able readers.

Summarization has a large number of studies (18) that replicate treatment effects, mainly at grades 5 and 6. Summarization presupposes writing as well as reading skill, hence its late study. The effects are largely specific to improving the writing of summaries, but there are 11 studies that show transfer effects on recall of what was summarized and on question answering. Standardized tests as general transfer were used rarely (only two studies). Instruction of summarization succeeds in that readers improve on the quality of their summaries of text, mainly identifying the main idea but also in leaving out detail, including ideas related to the main idea, generalizing, and removing redundancy. This indicates that summarizing is a good method of integrating ideas and generalizing from the text information. Furthermore, the instruction of summarization improves memory for what is read, both in terms of free recall and answering questions. This strategy instruction is used as a part of treatments that teach multiple strategies.

Multiple Strategy Instruction represents an evolution in the field from the study of individual strategies to their flexible and multiple use. This method finds considerable scientific support for its effectiveness as a treatment, and it is the most promising for use in classroom instruction where teachers and readers interact over texts. The NRP reviewed 11 studies not covered by the meta-analysis of Rosenshine and Meister (1994), who reviewed 16 reciprocal teaching studies on readers in grades 3 through 7.

One of the main methods is to have the teacher model an approach by showing how she or he would try to understand the text, using two or more combinations of four strategies: question generation, summarization, clarification, and prediction of what might occur. Rosenshine and Meister found strong evidence that the reciprocal teaching treatment showed near transfer. Experiment tests in ten studies had an average effect size of 0.88. There was also support for general transfer in nine studies where the average effect size was 0.32. All readers show more near transfer benefit in these treatments, whereas only the better readers show significant effect sizes in the 0.32 range. These data suggest that good readers benefit and generalize what they learn as strategies more than poor or below-average readers. Furthermore, the significant effect sizes do not occur for grade 3, are mixed for grades 4 through 6, and do occur for grades 7 and 8.

There were 11 other multiple strategy studies on readers in grades 2 through 11, with grade 4 as the modal grade. The strategies taught varied across these studies. In 6 of the 12 studies, students were taught summarizing or identification of main ideas. Three studies used question answering or generation, two used monitoring, and others used cooperative reading, recall, retelling, hypothesis testing, story structure, and psycholinguistic training (word, phrase, and sentence classification, morphological analysis). There was evidence for specific learning and near transfer. No studies reported the use of standardized tests.

Taken together, the evidence supports the use of combinations of reading strategies in natural learning situations. These findings build on the empirical validation of strategies alone and attest to their use in the classroom context. A common aspect of individual and multiple strategy instruction is the active involvement of motivated readers who read more text as a result of the instruction. These motivational and reading practice effects may be important to the success of multiple strategy instruction. Furthermore, multiple strategy instruction that is flexible as to which strategies are used and when they are taught over the course of a reading session provides a natural basis on which teachers and readers can interact over texts.

Discussion

In the preceding section, the Panel summarized the research claims and implications for instruction of comprehension. In this section, the kinds of claims being made are illustrated by three quotations:

> "The best way to pursue meaning is through conscious, controlled use of strategies" (Duffy, 1993, p. 223).
>
> "Becoming an effective transactional strategies instruction teacher takes several years" (Brown et al., 1996, p. 20).
>
> "The data suggests that students at all skill levels would benefit from being taught these strategies" (Rosenshine, Meister, & Chapman, 1996, p. 201).

The past 2 decades of research appear to support the enthusiastic advocacy of instruction of reading strategies expressed in the above quotations. The Panel's review of the literature indicates that there has been an extensive effort to identify reading comprehension strategies that can be taught to students to increase their comprehension and memory for text. The instruction of cognitive strategies improves reading comprehension in readers with a range of abilities.

This improvement occurs when teachers demonstrate, explain, model, and implement interaction with students in teaching them how to comprehend a text. In studies involving even a few hours of preparation, instructors taught students who were poor readers but adequate decoders to apply various strategies to expository texts in reading groups, with a teacher demonstrating, guiding, or modeling the strategies, and with teacher scaffolding (e.g., Palinscar & Brown, 1984; see Rosenshine, Meister, & Chapman, 1996 for a review). Such instruction is consistent with socially mediated learning theory (Pressley & McCormick, 1995; Vygotsky, 1978). Students using these strategies, even in limited ways, produced noticeable improvement in the use of the instructed strategies, albeit with only modest improvement on standardized reading tests (Rosenshine & Meister, 1994). More intensive instruction and modeling have been more successful in improving reading and standardized test scores (Bereiter & Bird, 1985; Block, 1993; Brown et al., 1996).

Many of the studies involve teaching one group of students a particular cognitive strategy to use while reading. These studies show that readers can learn a strategy and use it effectively in improving their comprehension. Reading, however, requires the coordinated and flexible use of several different kinds of strategies. Considerable success has been found in improving comprehension by instructing students on the use of more than one strategy during the course of reading. Skilled reading involves an ongoing adaptation of multiple cognitive processes. Becoming an independent, self-regulated, thinking reader is a goal that can be achieved through instruction of text comprehension (Brown et al., 1996).

Rosenshine and Meister (1994) conclude that the main weakness in understanding the practice of instruction is that not enough studies have been devoted to implementation. The NRP concurs with this conclusion.

Implementation of Instruction in Reading Comprehension

The major problem facing the teaching of reading comprehension strategies is that of implementation in the classroom by teachers in a natural reading context with readers of various levels on reading materials in content areas. For teachers, the art of instruction involves a series of "wh" questions: knowing *when* to apply *what* strategy with *which* particular student(s). Having students actually develop independent, integrated strategic reading abilities may require subtle instructional distinctions that go well beyond techniques such as instruction, explanation, or reciprocal teaching (Duffy, 1993). Duffy argues that strategies are not skills that can be taught by drill; they are plans for constructing meaning. Teaching students to acquire and use strategies may require altering traditional approaches to strategy instruction. It may be necessary to free teachers of the expectation that their job is to follow directions narrowly. Being strategic is much more than knowing the individual strategies. When faced with a comprehension problem, a good strategy user will coordinate strategies and shift strategies as it is appropriate to do so. They will constantly alter, adjust, modify, and test until they construct meaning and the problem is solved.

How well has the knowledge gleaned from research filtered into the classroom to impact teachers' actual practice? In spite of apparent effectiveness, teachers may not be using effective comprehension instruction strategies without having themselves had preparation in instruction (Anderson, 1992; Bramlett, 1994; Brown, 1996; Duffy, 1993; Durkin, 1979; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989; Pressley, 1998; Reutzel and Cooter, 1988). Pressley (1998) reports that a yearlong observation of ten upstate New York grade 4 and 5 classes in the 1995–1996 school year showed that teachers varied in several factors: their class management, their extent of monitoring student progress, their extent of engaging students, how concerned they were with external standards and state tests, and their frequency of assigning homework and skills practices. However, regarding comprehension instruction:

> In some classrooms . . . we observed explicit comprehension instruction only rarely, despite a great deal of research in the past two decades on how to promote children's comprehension of what they read . . . Indeed, the situation seemed to be much as Durkin (1979) described it two decades ago, with a great deal of testing of comprehension but very little teaching of it (Pressley, 1998, p. 198).

Durkin (1981) observed that when comprehension skill instruction is present, in many classrooms teachers appear to be "mentioning" a skill to students and "assigning" it to them rather than employing the effective instruction modeling and transactional practices that research supports (Durkin, 1981; Reutzel & Cotter, 1988). In the United States, reading from basal reading series accounts for 75% to 90% of classroom reading instruction time (Franklin et al., 1992). Although some basal teachers' manuals do provide more evaluative comprehension skill lessons, these lessons are usually not instructional and offer little structure and rationale for helping teachers give effective skill instruction (Reutzel & Cotter, 1988).

In a 5-year study of how teachers help low-achieving students become strategic readers, using monthly inservice strategy preparation sessions, biweekly individual teacher coaching with a strategy expert staff developer, and collaborative discussion of principals' and teachers' experiences in individual schools, Duffy (1993) suggests that effective reading instruction is associated more with independent teacher action than with implementation of basal text prescriptions. He argues that developing metacognitive readers who understand their reasoning requires teachers who themselves understand their reasoning, as well as a supportive environment in the schools for strategy learning. Pressley's (1998) recent observations suggest that too little has changed in the classroom since Durkin's 1978–1979 school year observations:

> A twist on this [1995–1996 school year] situation, however, was that the comprehension tasks now being given to students did seem to be informed by the comprehension process research of the past two decades. It was not uncommon, for example, for students to be asked to respond to short-answer questions requiring them to summarize what they read, identify confusing points in a text, construct questions pertaining to a text, or predict what might be next in a text. That is, they were asked to respond to questions constructed around the cognitive processes involved in skilled comprehension (i.e., summarizing, monitoring confusion, self-questioning, predicting based on prior knowledge). However, there was little evidence that students were being taught to self-regulate comprehension processes as they read, and in some classrooms, there was no evidence that they were being taught the active comprehension process validated in the last two decades. In general, students were provided with opportunities to practice comprehension strategies, but were not actually taught the strategies themselves nor the utility value of applying them. (Pressley, 1998, p. 198).

Deshler and Schumaker (1988) have taught learning disabled students how to comprehend, write, and remember in a learning disabilities curriculum. They emphasize the role of controllable factors, such as the use of strategies. One problem they encountered is that learning disabled students make attributions that render them dysfunctional (e.g., "I am stupid."). These kinds of attributions can defeat what might otherwise be effective comprehension instruction. Alternatively, effective comprehension instruction might lead learning disabled students to make more positive, functional attributions.

When conscientious, diligent, and highly professional teachers apply their strategy instruction in the classroom, even when applied imperfectly, their students do improve in reading comprehension (Bramlett, 1994; Duffy, 1993; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989). However, close observation of inservice trained strategy teachers suggests that:

> Progress was not easily accomplished. It was a struggle. For much of the academic year, the four [strategic] teachers [in the study] required from their students counterproductive 'answers' and 'routes'—that is, answers and thinking that led students to construct inaccurate conceptions [of strategies].

> Although by May it appeared that [their grade 2 poor reading] students were developing an integrated concept of what it means to be strategic, students' responses to interview probes during fall and winter suggested incomplete conceptions or *mis*conceptions about what it means to be strategic (Duffy, 1993, p. 237).

In spite of heavy emphasis on modeling and metacognitive instruction, even very good teachers may have trouble implementing, and may even omit, crucial aspects of strategic reasoning. The research suggests that, when partially implemented, students of strategy teachers will still improve. But it is not easy for teachers or readers to develop readers' conceptions about what it means to be strategic. It takes time and ongoing monitoring of success to evolve readers into becoming good strategy users.

> Helping teachers [become good strategy teachers] will require a significant change in how teacher educators and staff developers work with teachers and what they count as important about learning to be a teacher. Current practices that require teachers to successfully complete university course work, to attend mandated half-day in-service programs, or to be 'trained' in the 'right way' to teach and then [be] held accountable for that encourage teachers, like the children . . . to learn only the labels of professional knowledge without learning how to be strategic themselves. Such practices must be replaced by teacher education/staff development experiences that account for (1) the complexity involved in teaching [students] to be strategic and for (2) the creative adaptations teachers must make as they deal with that complexity (Duffy, 1993, p. 244-245).

Strategic reading requires strategic teaching, which involves putting teachers in positions where their minds are the most valued educational resource (Duffy, 1993). Skilled reading is constructive reading, and the activities of the reader matter (Pressley, Harris, & Marks, 1992; Pressley & Afflerbach, 1995).

What is the scientific basis for claims made about instruction of comprehension?

The Panel now begins a more critical analysis of the literature on instruction of comprehension. First, the quality of the studies is discussed. Second, scientific criteria are applied and the Panel's prior evaluations to arrive at an overall set of conclusions are discussed.

Quality of Studies: An Overlooked Issue

In half the studies reviewed by Rosenshine and Meister (1994), experimenters failed to address the quality of instruction in the intervention study. There are several papers, however, that have raised questions about the quality issues of reading research: Almasi, Palmer, Gambrell, and Pressley (1994); Lysynchuk, Pressley, d'Ailly, Smith, and Cake (1989); Pressley et al. (1989); Rosenshine et al. (1996); Rosenshine and Meister (1994); and Troia (1999). Of these, Lysynchuk et al. (1989) evaluated the methodological adequacy of 37 studies of reading comprehension instruction. Several problems were identified. Of particular importance were (1) failure to randomly assign students to treatments and control conditions, (2) failure to expose experimental and control participants to the same training materials, (3) failure to provide information about the amount of time spent on dependent variable tasks, (4) failure to study fidelity of treatment by not including analysis of teacher and reader performance during instruction, (5) use of inappropriate units (individual, group, classroom) in analyses, and (6) failure to assess either long-term effects or generalization of the strategies to other tasks and materials.

Lysynchuk et al. (1989) applied 24 criteria of internal validity (classified in four categories as to general design, possible confounds, measurement, and statistics) and five criteria of external validity (theory, sample, reading ability, text properties, measures of transfer). The range of percentages of studies that met internal validity criteria was from 17 to 100, median = 78%. For external validity, the range was from 8% to 100%, median = 82.5 percent. Although most studies specified the experimental and control groups and the independent and dependent variables in their general design presentations, only 64% randomly assigned participants or classes to the experimental and control conditions, compromising cause-and-effect conclusions.

With respect to confounds, in 75% of the studies, control subjects were lead to believe that they were in an experimental condition; therefore, 25% were not, allowing for possible Hawthorne effects. In nearly one-third of the studies, there were possible confounds of differences in training materials between the experimental and control groups with the experimental groups given more materials to read. However, in these studies they were, with one exception, exposed to materials for the same amount of time.

In other studies, time on task was confounded with condition. Experimental groups may have been allowed more time to read than control groups. Only 10 of 37 studies reported the amount of time, and 8 of 10 of these were the same. However, these studies did not analyze what students did during the time assigned; therefore, it is unknown whether they used the time to read. In addition, there were possible experimenter-by-condition or teacher-by-condition confounds in some studies because neither the experimenters nor the teachers were randomly assigned to groups. Measurement problems involved not measuring reliability (37% of the studies), floor and ceiling effects (33% of the studies), and failure to assess fidelity of treatment through checks on manipulation (only 37% did so for teachers, and 27% measured ongoing processes). On statistical practices, the most serious flaw was in the use of appropriate units—if one assigns groups to conditions and then conducts analyses on individuals, the unit of analysis differs from the unit of treatment. Errors then cannot be assumed to be independent. With respect to external validity, most studies met theory and reporting of sample criteria. Other problems involved omission of data on reading level (16%), failures to measure transfer or delayed effects (76%), and failures to measure transfer to school subjects (92%).

Future studies would benefit from attention to quality criteria for internal and external validity. In particular, researchers should conduct reliability assessments of their scoring of data when raters are used; should use random assignment of experimenters, teachers, classrooms, or students where possible; or should at least collect data on comparability of instructors and on participant characteristics in the treatment and control conditions. Researchers should try to meet quasi-experimental criteria if random assignment is not possible (Cook & Cambell, 1979). Hawthorne effects can be reduced by motivating controls to believe that they are receiving the same benefits and treatment as experimental participants. Often the tasks themselves motivate experimental and controls differently, confounding motivation with the variable of study. Similarly, Hawthorne effects on teachers can occur if they believe that the experimental group will benefit more than controls. One way to deal with this problem is to assign the teacher to both groups but with the belief that either treatment would benefit the participants.

Future studies should include fidelity to treatment measures of the preparation of teachers, of the teachers' teaching the strategies as intended, and of the students' performance during training. There is a need to observe, document, and analyze all components of the experiment, from training to implementation to learning to assessment. The amount of time on each task should be recorded and reported as well as examined in relation to outcome measures. Floor and ceiling effects on measures should be avoided. The unit of analysis should be the same as the unit of treatment. All these steps would improve the design and internal validity of studies on reading strategy instruction. External validity could be improved by the inclusion and measurement of training and transfer of training to other measures, particularly performance in content areas. Text, as a variable, has been sorely neglected. The external validity of a study could also be improved by the kind of texts used (both expository and narrative and sampled from content areas), an analysis of text difficulty, the content and structure of the text, the vocabulary and sentence complexity of the text, appropriateness of the level of text difficulty to the ability of readers, and possible interactions between difficulty of the text and ability of reader. Long-term benefits could be assessed through followup studies later so that the effects are not just short term.

In the section of this Text Comprehension report on quality of studies, the Panel describes a set of criteria for internal and external validity that should be used to plan, conduct, and report research in individual studies but also that can be applied in evaluation of single and multiple studies and reviews of studies. That section includes several criteria for internal and external validity. These criteria incorporate, elaborate, extend, and adapt to the reading situation the 24 categories of the Lysynchuk et al. (1989) review.

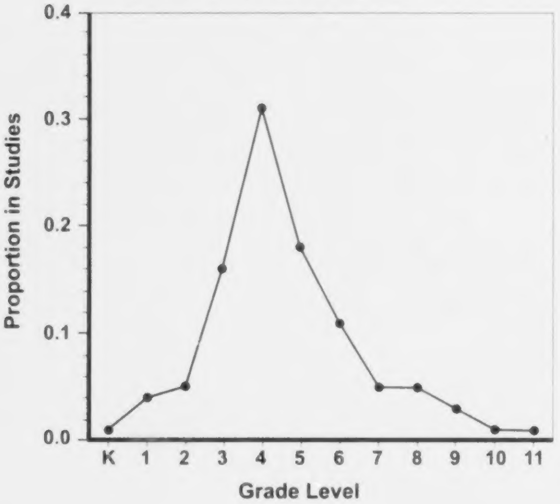
Scientific Evaluation of the Claims Made in the Literature

The empirical evidence reviewed favors the conclusion that teaching of a variety of reading comprehension strategies leads to increased learning of the strategies, to specific transfer of learning, to increased memory and understanding of new passages, and, in some cases, to general improvements in comprehension. In particular, individual strategies that can be used in natural reading or content area instruction and through interaction with the teacher over a text appear to have a strong scientific support for their effectiveness and for their inclusion in classroom programs on comprehension instruction.

The NRP now integrates its evaluations of the instruction strategies that have the best scientific basis for effectiveness and use by teachers in the classroom. The Panel first considers the grade level appropriateness and general effectiveness, then the evidence of reliability, robustness, replication, and transfer for a set of particular strategies in support of the general conclusion above.

On what grade levels has text comprehension instruction been effectively studied? Figure 1 shows the distribution of the grade levels at which investigations of instruction in comprehension have been successfully carried out.

Figure 1. Distribution of Grade Levels Across All Studies of Direction



In Figure 1, grades 3 through 6 constitute 76% of the grade levels studied. The modal grade is 4 with the next highest percentages occurring with grades 3 and 5. Thus, instruction of comprehension begins mainly at the 3rd grade and continues through the 6th grade. In examining the studies, the Panel found that the lower three grades (K through 2) were studied primarily as a part of an experimental curriculum. The higher grades (above grade level 6) tend to focus on less able readers. The increase in percentage at grade level 3 suggests that researchers taught readers who had achieved decoding and other basic reading skills before they were taught strategies.

To determine the effectiveness of instruction and whether it was related to grade level, the Panel found the percentage of reported significant findings where the experimental treatment was favored over the control group. The overall average percentages of success, as measured by experimenter tasks or by standardized tests, were 97 and 93%, respectively. The high overall rates of success are not surprising because these data are based upon published studies. For grades K through 1 and 7 through 11, the reported percentage of success was 100 on experimenter tasks and standardized tests; for grades 2 through 6, the average was 92%. For standardized tests, the average success was 89% for grades 2 through 6. There was no relationship between grade level and the respective percentages of success in treatment.

These data indicate that instruction is likely to be more successful when measured on experimenter designed tasks than on standardized tests of comprehension. The instruction of comprehension appears to be effective on grades 3 through 6.

With respect to the scientific basis of the instruction of text comprehension, the NRP concludes that comprehension instruction can effectively motivate and teach normal readers to learn and to use comprehension strategies that benefit them.

These comprehension strategies yield increases in measures of near transfer such as recall, question answering and generation, and summarization of texts. Furthermore, when used in combination, these comprehension strategies produce general gains on standardized comprehension tests. Teachers can learn to teach students to use comprehension strategies in natural learning situations. In addition, when teachers teach these strategies, their students learn them and improve their reading comprehension.

A common aspect of individual and multiple strategy instruction is the active involvement of motivated readers who read more text as a result of the instruction. These motivational and reading practice effects may be important to the success of multiple strategy instruction.

Multiple strategy instruction that is flexible as to which strategies are used and when they are taught over the course of a reading session provides a natural basis on which teachers and readers can interact over texts. The research literature developed from early studies of isolated strategies then moved to the use of strategies in combination, and finally to the preparation of teachers to teach strategies in interactions about texts with readers in naturalistic settings. The Panel regards this development as the most important finding of its review because it moves from the laboratory to the classroom and prepares teachers to teach strategies in ways that are effective and natural.

The empirical evidence reviewed favors the conclusion that teaching of a variety of reading comprehension strategies leads to increased learning of the strategies, to specific transfer of learning, to increased memory and understanding of new passages, and, in some cases, to general improvements in comprehension.

The important development of instruction of comprehension research is the study of teacher preparation for instruction of multiple, flexible strategies with readers in natural settings and content areas and the assessment of the effectiveness of this instruction by prepared teachers on comprehension.

Directions for Further Research

The Panel's analysis of the research on instruction of text comprehension left a number of questions unanswered:

1. More information is needed on the effective ways to teach teachers how to use proven strategies for instruction in text comprehension. This information is crucial to situations where teachers and readers interact over texts in real classroom contexts.
2. The Panel reviewed some evidence that instruction in comprehension in content areas benefit readers in terms of achievement in social studies. There is a need to know whether instruction of comprehension strategies leads to learning skills that improve performance in content areas of instruction. If so, it might be efficient to teach reading comprehension as a learning skill in content areas.
3. It is already known that instruction of comprehension has been successful over the grade 3 through 6 range. Further evidence is needed on whether certain strategies are more appropriate for certain ages and abilities, what the important reader characteristics are that influence successful instruction of reading comprehension, and which strategies, in combination, are best for younger readers, poor or below-average readers, and for learning disabled and dyslexic readers.
4. It is also important to know whether successful instruction generalizes across different text genres (e.g., narrative and expository) and across texts from different subject content areas. The NRP's review of the research indicated that little or no attention has been given to the kinds of text used. The review also indicated that there was little available information on the difficulty level of texts.
5. Information is needed on the important teacher characteristics that influence successful instruction of reading comprehension, as well as the effective ways to prepare teachers, both preservice and inservice.

6. Prior studies suffer when the quality of the studies is assessed (Lysynchuk et al., 1989) according to criteria of internal and external validity. These issues need to be considered when designing future research. The main problems were:

   (a) Failure to randomly assign students to treatments and control conditions and failure to expose experimental and control participants to the same training materials

   (b) Failure to provide information about the amount of time spent on dependent variable tasks

   (c) Failure to study fidelity of treatment, by failing to analyze teacher and reader performance during instruction

   (d) Use of inappropriate units (individual, group, classroom) in analyses

   (e) Failure to assess either long-term effects or generalization of the strategies to other tasks and materials.

References

The references of this report are listed, first, as references cited in text, and second, as references used in each category of text comprehension instruction.

Text References

Almasi, J. F., Palmer, B. M., Gambrell, L. B., & Pressley, M. (1994). Toward disciplined inquiry: A methodological analysis of whole-language research. Educational Psychologist, 29, 193-202

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson (Ed.), Handbook of reading research (pp. 255-291). New York: Longman.

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. Teaching and Teacher Education, 8(4), 391-403.

Anderson, V., & Riot, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in grades 6-20. Special Issue: Strategies instruction. Elementary School Journal, 94(2), 121-137.

Athey, I. (1983). Language Development factors related to reading development. Journal of Educational Research, 76(4), 197-203.

Baumann, J. F. (1986). Teaching third-grade students to comprehend anaphoric relationships: The application of a instruction model. Reading Research Quarterly, 21(1), 70-90.

Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. Journal of Reading Behavior, 24(2), 143-172.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. Cognition and Instruction, 2, 131-156.

Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. Journal of School Psychology, 32(1), 67-84.

Borduin, B. J., Borduin, C. M., & Manley, C. M. (1994). The use of imagery training to improve reading comprehension of second graders. Journal of Genetic Psychology, 155(1), 115-118.

Capelli, C. A. & Markman, E.M. (1982). Suggestions for training comprehension monitoring. Topics in Learning & Learning Disabilities, 2, 87-96.

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design and analysis issues for field settings. Chicago: Rand McNally College Publishing Company.

Deshler, D. D., & Schumaker, J. B. (1988). An instructional model for teaching students how to learn. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.). Alternative educational delivery systems: Enhancing instructional options for all students (pp. 391-411). Washington, DC: National Association of School Psychologists.

Dickson, W. P. (Ed.) (1981). Children's oral communication skills. New York: Academic Press.

Duffy, G., Lanier, J. E., & Roehler, L. R. (1980). On the need to consider instructional practice when looking for instructional implications. Paper presented at the Reading Expository Materials, University of Wisconsin-Madison.

Duffy, G. G., & Roehler, L. R. (1989). Why strategy instruction is so difficult and what we need to do about it. In C. B. McCormick, G. Miller, & M. Pressley (Eds.), Cognitive strategy research: From basic research to educational applications. New York: Springer-Verlag.

Durkin, D. (1979). What classroom observations reveal about reading comprehension. Reading Research Quarterly, 14, 518-544.

Durkin, D. (1981). Reading comprehension instruction in five basal reading series. Reading Research Quarterly, 14, 481-533.

Durkin, D. (1993). Teaching them to read (6th ed.). Boston, MA: Allyn & Bacon.

Englert, C. S., & Raphael, T. (1989). Constructing well-formed prose: Process, structure, and metacognitive knowledge. Exceptional Children, 54, 513-520.

Franklin, M. R., Roach, P. B., & Clary, E., Jr. (1992). Overcoming the reading comprehension barriers of expository texts. Educational Research Quarterly, 16 (1), 5-14.

Gagne, R. M. (1977). The conditions of learning (3rd ed). New York: CBS College Publishing.

Grant, J., Elias, G., & Broerse, J. (1989). An application of Palinscar and Brown's comprehension instruction paradigm to listening. Contemporary Educational Psychology, 14(2), 164-172.

Harris, T. L., & Hodges, R. E. (Eds.) (1995). The literacy dictionary: The vocabulary of reading and writing. Newark, DE: International Reading Association.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of discourse comprehension and production. Psychological Review, 83, 363-394.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. Elementary School Journal, 99(1), 3-22.

Larkin, J. H., & Reif, F. (1976). Analysis and teaching of a general skill for studying scientific text. Journal of Educational Psychology, 72, 348-350.

Levin, J. R., & Divine-Hawkins, P. (1974). Visual imagery as a prose-learning process. Journal of Reading Behavior, 6, 23-30.

Lysynchuk, L. M., Pressley, M., D'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. Reading Research Quarterly, 24(4), 458-470.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. Elementary School Journal, 90(5), 469-484.

Markman, E. M. (1978). Realizing that you don't understand: A preliminary investigation. Child Development, 48, 986-992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50(3), 643-655.

Markman, E. M. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), Children's oral communication skills (pp. 61-84). New York: Academic Press.

Mier, M. (1984). Comprehension monitoring in the elementary classroom. Reading Teacher, 37(8), 770-774.

Palinscar, A. S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 2, 117-175.

Paris, S. G., Saarnio, D. A., & Cross, D. R. (1986). A metacognitive curriculum to promote children's reading and learning. Australian Journal of Psychology, 38(2), 107-123.

Peters, E. E., & Levin, J. R. (1986). Effects of a mnemonic imagery strategy on good and poor readers' prose recall. Reading Research Quarterly, 21, 179-192.

Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. NY: The Guilford Press.

Pressley, M., & Afflerbach, P. (1995). Verbal protocols of reading: The nature of constructively responsive reading. Mahwah, NJ: Erlbaum.

Pressley, M., Almasi, J., Schuder, T., Bergman, J., & Kurita, J. A. (1994). Transactional instruction of comprehension strategies: The Montgomery County, Maryland, SAIL program. Reading & Writing Quarterly: Overcoming Learning Difficulties, 10(1), 5-19.

Pressley, M., Gaskins, I. W., Wile, D., Cunicelli, E. A., & Sheridan, J. (1991). Teaching literacy strategies across the curriculum: A case study at Benchmark School. In J. Zutell & S. McCormick (Eds.). Learner factors/teacher factors: Issues in liberacy research and instruction: Fortieth Yearbook of the National Reading Conference (pp. 219-228). Chicago, IL: The National Reading Conference.

Pressley, M., Harris, K. R..,& Marks, M. B. (1992). But good strategy instructors are constructivists! Educational Psychology Review, 4(1), 3-31.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J. A., & Kurita, J. A.. (1989). Strategies that improve children's memory and comprehension of text. Elementary School Journal, 90(1), 3-32.

Pressley, M., & McCormick, C. B. (1995). Cognition, teaching, and assessment. NY: Harper Collins College Publishers.

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. Reading Research Quarterly, 21(4), 422-438.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. Review of Educational Research, 64(4), 479-530.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. Review of Educational Research, 66(2), 181-221.

Reutzel, D. R., & Cooter, R. B., Jr. (1988). Research implications for improving basal skill instruction. Reading Horizons, 28(3), 208-215.

Scardamalia, M., & Beretier, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S. F. Chipman, J.W. Segal, & R. Glaser (Eds.). Thinking and learning skills: Vol. 2. Research and open questions (pp. 563-577). Hillsdale, NJ: Erlbaum.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. C. Wittrock (Ed.), Handbook of research on teaching, 3rd ed. (pp. 778-803). NY: MacMillan.

Schoenfeld, A. (1985). Mathematical problem solving. NY: Academic Press.

Spires, H. A., Gallini, J., & Riggsbee, J. (1992). Effects of schema-based and text structure-based cues on expository prose comprehension in fourth graders. Journal of Experimental Education, 60(4), 307-320.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and instruction in reading comprehension strategies on main idea identification. Journal of Educational Psychology, 83(1), 8-16.

Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository material. Journal of Educational Psychology, 74(3), 323-340.

Troia, G. A. (1999) Phonological awareness intervention research: A critical review of the experimental methodology. Reading Research Quarterly, 34, 28-52.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes In M. Cole, V. John-Steiner, S. Scriber, and E. Souberman (Eds. & Trans.). Cambridge, MA: Harvard University Press.

Weinstein, C. E. (1978). Teaching cognitive elaboration learning strategies. In H. F. O'Neal (Ed.). Learning strategies. NY: Academic Press.

Whipple, G. (Ed.). (1925). The Twenty-fourth Yearbook of the National Society for the Study of Education: Report of the National Committee on Reading. Bloomington, IL: Public School Publishing Company.

Wright, B., & Stenner, A. J. (1998). Readability and reading ability. Paper presented at the Australian Council on Education Research (ACER), Melbourne, Australia.

Category References

Comprehension Monitoring

Babbs, P. J. (1984). Monitoring cards help improve comprehension. Reading Teacher, 38(2), 200-204.

Baker, L., & Zimlin, L. (1989). Instructional effects on children's use of two levels of standards for evaluating their comprehension. Journal of Educational Psychology, 81(3), 340-346.

Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. Journal of Reading Behavior, 24(2), 143-172.

Block, C. C. (1993). Strategy instruction in a literature-based reading program. Special issue: Strategies instruction. Elementary School Journal, 94(2), 139-151.

Carr, E. M., Dewitz, P., & Patberg, J. P. (1983). The effect of inference training on children's comprehension of expository text. Journal of Reading Behavior, 15(3), 1-18.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. Journal of Educational Psychology, 80(2), 131-142.

Elliot-Faust, D. J., & Pressley, M. (1986). How to teach comparison processing to increase children's short- and long-term listening comprehension monitoring. Journal of Educational Psychology, 78, 27-33.

Hasselhorn, M., & Koerkel, J. (1986). Metacognitive versus traditional reading instructions: The mediating role of domain-specific knowledge on children's text-processing. Human Learning: Journal of Practical Research and Applications, 5(2), 75-90.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. Child Development, 46, 986-992.

Miller, G. E. (1985). The effects of general and specific self-instruction training on children's comprehension monitoring performances during reading. Reading Research Quarterly, 20(5), 616-628.

Miller, G. E. (1987). The influence of self-instruction on the comprehension monitoring performance of average and above average readers. Journal of Reading Behavior, 19(3), 303-317.

Miller, G. E., Giovenco, A., & Rentiers, K. A. (1987). Fostering comprehension monitoring in below average readers through self-instruction training. Journal of Reading Behavior, 19(4), 379-394.

Nelson, C. S., et al. (1996). The effect of teacher scaffolding and student comprehension monitoring on a multimedia/interactive videodisc science lesson for second graders. Journal of Educational Multimedia and Hypermedia, 5(3-4), 317-348.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. Journal of Educational Psychology, 76(6), 1239-1252.

Paris, S. G., & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. Child Development, 55(6), 2083-2093.

Paris, S. G., Saarnio, D. A., & Cross, D. R. (1986). A metacognitive curriculum to promote children's reading and learning. Australian Journal of Psychology, 38(2), 107-123.

Payne, B. D., & Manning, B. H. (1992). Basal reader instruction: Effects of comprehension monitoring training on reading comprehension, strategy use and attitude. Reading Research and Instruction, 32(1), 29-38.

Schmitt, M. C. (1988). The effects of an elaborated directed reading activity on the metacomprehension skills of third graders. National Reading Conference Yearbook, 37, 167-181.

Schunk, D. H., & Rice, J. M. (1984). Strategy self-verbalization during remedial listening comprehension instruction. Journal of Experimental Education, 53(1), 49-54.

Schunk, D. H., & Rice, J. M. (1985). Verbalization of comprehension strategies: Effects on children's achievement outcomes. Human Learning: Journal of Practical Research and Applications, 4(1), 1-10.

Silven, M. (1992). The role of metacognition in reading instruction. Scandinavian Journal of Educational Research, 36(3), 211-221.

Tregaskes, M. R., & Daines, D. (1989). Effects of metacognitive strategies on reading comprehension. Reading Research and Instruction, 29(1), 52-60.

Cooperative Learning References

Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. Journal of School Psychology, 32(1), 67-84.

Guthrie, J. T., et al. (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. Reading Research Quarterly, 31(3), 306-332.

Judy, J. E., Alexander, P. A., Kulikowich, J. M., & Wilson, V. L. (1988). Effects of two instructional approaches and peer tutoring on gifted and non gifted sixth-grade students' analogy performance. Reading Research Quarterly, 23(2), 236-256.

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. Elementary School Journal, 99(1), 3-22.

Mathes, P. G., et al. (1994). Increasing strategic reading practice with Peabody classwide peer tutoring. Learning Disabilities Research and Practice, 9(1), 44-48.

Pickens, J., & McNaughton, S. (1988). Peer tutoring of comprehension strategies. Educational Psychology: An International Journal of Experimental Educational Psychology, 8(1-2), 67-80.

Soriano, M., Vidal-Abarca, E., & Miranda, A. (1996). Comparación de dos procedimentos de instrucción en comprensión y aprendizaje de textos: Instrucción directa y enseñanza recíproca. [Comparison of two procedures for instruction in comprehension and text learning: Direct instruction and reciprocal teaching.] Infancia y Aprendizaje, 74, 57-65.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. Reading Research Quarterly, 22(4), 433-454.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and instruction in reading comprehension strategies on main idea identification. Journal of Educational Psychology, 83(1), 8-16.

Uttero, D. A. (1988). Activating comprehension through cooperative learning. Reading Teacher, 41(4), 390-395.

Curriculum Plus Strategies References

Au, K. H., & Scheu, J. A. (1989). Guiding students to interpret a novel. Reading Teacher, 43(2), 104-110.

Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. Journal of School Psychology, 32(1), 67-84.

Feathers, K. M., & Smith, F. R. (1987). Meeting the reading demands of the real world: Literacy based content instruction. Journal of Reading, 30(6), 506-511.

Freeman, R. H., & Freeman, G. G. (1987). Reading acquisition: A comparison of four approaches to reading instruction. Reading Psychology, 8(4), 257-272.

Noyce, R. M., & Christie, J. F. (1983). Effects of an integrated approach to grammar instruction on third graders' reading and writing. Elementary School Journal, 84(1), 63-69.

Romance, N. R., & Vitale, M. R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. Journal of Research in Science Teaching, 29(6), 545-554.

Tharp, R. G. (1982). The effective instruction of comprehension: Results and description of the Kamehameha early education program. Reading Research Quarterly, 17(4), 503-527.

Webster, R. E., & Braswell, L. A. (1991). Curriculum bias and reading achievement test performance. Psychology in the Schools, 28(3), 193-199.

Graphic Organizer References

Alvermann, D. E., & Boothby, P. R. (1983). A preliminary investigation of the differences in children's retention of "inconsiderate" text. Reading Psychology, 4(3-4), 237-246.

Alvermann, D. E., & Boothby, P. R. (1986). Children's transfer of graphic organizer instruction. Reading Psychology, 7(2), 87-100.

Armbruster, B. B., Anderson, T. H., & Meyer, J. L. (1991). Improving content-area reading using instructional graphics. Reading Research Quarterly, 26(4), 393-416.

Armbruster, B. B., Anderson, T. H., & Meyer, J. L. (1992). "Improving content-area reading using instructional graphics": Erratum. Reading Research Quarterly, 27(3), 282.

Baumann, J. F. (1984). The effectiveness of an instruction paradigm for teaching main idea comprehension. Reading Research Quarterly, 20(1), 93-115.

Berkowitz, S. J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. Reading Research Quarterly, 21(2), 161-178.

Darch, C. B., Carnine, D. W., & Kameenui, E. J. (1986). The role of graphic organizers and social structure in content area instruction. Journal of Reading Behavior, 18(4), 275-295.

Gordon, C. J., & Rennie, B. J. (1987). Restructuring content schemata: An intervention study. Reading Research and Instruction, 26(3), 162-188.

Simmons, D. C., et al. (1988). Effects of teacher-constructed pre- and post-graphic organizaer instruction on sixth-grade science students' comprehension and recall. Journal of Educational Research, 82(1), 15-21.

Sinatra, R. C., Stahl-Gemake, J., & Berg, D. N. (1984). Improving reading comprehension of disabled readers through semantic mapping. Reading Teacher, 38(1), 22-29.

Vidal-Abarca, E., & Gilabert, R. (1995). Teaching strategies to create visual representations of key ideas in content area text materials: A long-term intervention inserted in school curriculum. Special Issue: Process-oriented instruction: Improving student learning. European Journal of Psychology of Education, 10(4), 433-447.

Listening References

Boodt, G. M. (1984). Critical listeners become critical readers in remedial reading class. Reading Teacher, 37(4), 390-394.

Shany, M. T., & Biemiller, A. (1995). Assisted reading practice: Effects on performance for poor readers in grades 3 and 4. Reading Research Quarterly, 30(3), 382-395.

Shepherd, T. R., & Svasti, S. (1987). Improved sixth grade social studies test scores via instruction in listening. Journal of Social Studies Research, 11(2), 20-23.

Sippola, A. E. (1988). The effects of three reading instruction techniques on the comprehension and word recognition of first graders grouped by ability. Reading Psychology, 9(1), 17-32.

Mental Imagery References

Borduin, B. J., Borduin, C. M., & Manley, C. M. (1994). The use of imagery training to improve reading comprehension of second graders. Journal of Genetic Psychology, 155(1), 115-118.

Gambrell, L. B., & Bales, R. J. (1986). Mental imagery and the comprehension-monitoring performance of fourth and fifth-grade poor readers. Reading Research Quarterly, 21, 454-464.

Peters, E. E., & Levin, J. R. (1986). Effects of a mnemonic imagery strategy on good and poor readers' prose recall. Reading Research Quarterly, 21, 179-192.

Pressley, G. M. (1976). Mental imagery helps eight-year olds remember what they read. Journal of Educational Psychology, 68, 355-359.

Shriberg, L. K., Levin, J. R., McCormick, C. B., & Pressley, M. (1982). Learning about "famous" people via the keyword method. Journal of Educational Psychology, 74, 238-247.

Mnemonics References

Levin, J. R., Levin, M. E., Glasman, L. D., & Nordwall, M. B. (1992). Mnemonic vocabulary instruction: Additional effectiveness evidence. Contemporary Educational Psychology, 17(2), 156-174.

Levin, J. R., Shriberg, L. D., & Berry, J. K. (1983). A concrete strategy for remembering abstract prose. American Educational Research Journal, 20, 277-290.

McCormick, C. B., & Levin, J. R. (1984). A comparison of different prose-learning variations of the mnemonic keyword method. American Educational Research Journal, 21, 379-398.

McCormick, C. B., Levin, J. R., Cykowski, F., & Danilovics, P. (1984). Mnemonic-strategy reduction of prose-learning interference. Educational Communication and Technology Journal, 32, 145-152.

Multiple Strategies References

Reciprocal Teaching Studies (Reviewed by Rosenshine & Meister, 1994)

Brady, P. I. (1990). Improving the reading comprehension of middle school students through reciprocal teaching and semantic mapping strategies. Unpublished doctoral dissertation, University of Alaska.

Dermody, M. (1988). Metacognitive strategies for development of reading comprehension for younger children. Paper presented at the American Association of Colleges for Teacher Education, New Orleans, LA.

Fischer Galbert, J. L. (1989). An experimental study of reciprocal teaching of expository test with third, fourth, and fifth grade students enrolled in chapter 1 reading. Unpublished doctoral dissertation, Ball State University, Muncie, IN.

Jones, M. P. (1987). Effects of reciprocal teaching method on third graders' decoding and comprehension abilities. Unpublished doctoral dissertation, Texas A&M University.

Labercane, G., & Battle, J. (1987). Cognitive processing strategies, self-esteem, and reading comprehension of learning disabled students. Journal of Special Education, 11, 167-185.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. Elementary School Journal, 90(5), 469-484.

Padron, Y. N. (1985). Utilizing cognitive reading strategies to improve English reading comprehension of Spanish-speaking bilingual students. Unpublished doctoral dissertation, University of Houston.

Palinscar, A. S. (1987). Collaborating for collaborative learning of text comprehension. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Palinscar, A. S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 2, 117-175.

Rich, R. Z. (1989). The effects of training adult poor readers to use text comprehension strategies. Unpublished doctoral dissertation, Columbia University, New York.

Rush, R. T., & Milburn, J.L. (1988). The effects of reciprocal teaching on self-regulation of reading comprehension in a post-secondary technical school program. Paper presented at the National Reading Conference, Tucson, AZ.

Shortland-Jones, B. (1986). The development and testing of an instructional strategy for improving reading comprehension based on schema and metacognitive theories. Unpublished doctoral dissertation. University of Oregon.

Taylor, B. M., & Frye, B. J. (1992). Comprehension strategy instruction in the intermediate grades. Reading Research and Instruction, 32(1), 39-48.

Williamson, R. A. (1989). The effect of reciprocal teaching on student performance gains in third grade basal reading instruction. Unpublished doctoral dissertation, Texas A&M University.

Other Reciprocal Teaching Studies (not reviewed by Rosenshine & Meister, 1994)

Carnine, D., & Kinder, D. (1985). Teaching low-performing students to apply generative and schema strategies to narrative and expository material. Remedial and Special Education, 6, 20-30.

Chan, L. D. S., & Cole, P. G. (1986). Effects of inference training on children's comprehension of expository text. Remedial and Special Education, 7, 33-40.

Gilroy, A., & Moore, D. W. (1988). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities with ten primary school girls. Special Issue: Changing academic behavior. Educational Psychology, 8(1-2), 41-49.

Grant, J., Elias, G., & Broerse, J. (1989). An application of Palinscar and Brown's comprehension instruction paradigm to listening. Contemporary Educational Psychology, 14(2), 164-172.

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. Educational Psychologist, 22, 255-278.

Kelly, M., Moore, D. W., & Tuck, B. F. (1994). Reciprocal teaching in a regular primary school classroom. Journal of Educational Research, 88(1), 53-61. Klingner, 1998

Klingner, J. K., Vaughn, S., & Schumm, J. S. (1998). Collaborative strategic reading during social studies in heterogeneous fourth-grade classrooms. Elementary School Journal, 99(1), 3-22.

Loranger, A. L. (1997). Comprehension strategies instruction: Does it make a difference? Reading Psychology, 18(1), 31-68.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. Elementary School Journal, 90(5), 469-484.

Palinscar, A. S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. Special issue: Cognitive instruction and problem learners. RASE: Remedial and Special Education, 12(3), 43-53.

Soriano, M., Vidal-Abarca, E., & Miranda, A. (1996). Comparación de dos procedimentos de instrucción en comprensión y aprendizaje de textos: Instrucción directa y enseñanza reciproca. [Comparison of two procedures for instruction in comprehension and text learning: Direct instruction and reciprocal teaching.] Infancia y Aprendizaje, 74, 57-65.

Taylor, B. M., & Frye, B. J. (1992). Comprehension strategy instruction in the intermediate grades. Reading Research and Instruction, 32(1), 39-48.

Other Multiple Strategy Treatments

Adams, A., Carnine, D., & Gersten, R. (1982). Instructional strategies for studying content area texts in the intermediate grades. Reading Research Quarterly, 18(1), 27-55.

Anderson, V., & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in grades 6-20. Special Issue: Strategies instruction. Elementary School Journal, 94(2), 121-137.

Blanchard, J. S. (1980). Preliminary investigation of transfer between single-word decoding ability and contextual reading comprehension by poor readers in grade six. Perceptual and Motor Skills, 51(3), t2.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. Journal of Educational Psychology, 88(1), 18-37.

Carr, E., Bigler, M., & Morningstar, C. (1991). The effects of the CVS strategy on children's learning.

Pelow, R. A., & Colvin, H. M. (1983). PQ4R as it affects comprehension of social studies reading material. Social Studies Journal, 12, 14-22 (Spring).

Reutzel, D. R., & Hollingsworth, P. M. (1991). Reading comprehension skills: Testing the distinctiveness hypothesis. Reading Research and Instruction, 30(2), 32-46.

Reutzel, D. R., & Hollingsworth, P. M. (1991). Reading time in school: Effect on fourth graders' performance on a criterion-referenced comprehension test. Journal of Educational Research, 84(3), 170-176.

Ritchie, P. (1985). Graduate research: reviews and commentary: The effects of instruction in main idea and question generation. Reading-Canada-Lecture, 3(2), 139-146.

Sindelar, P. T. (1982). The effects of cross-aged tutoring on the comprehension skills of remedial reading students. Journal of Special Education, 16(2), 199-206.

Smith, K., Johnson, D. W., & Johnson, R. T. (1981). Can conflict be constructive? Controversy versus concurrence seeking in learning groups. Journal of Educational Psychology, 73(5), 651-663.

Stevens, R. J. (1988). Effects of strategy training on the identification of the main idea of expository passages. Journal of Educational Psychology, 80(1), 21-26.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. Reading Research Quarterly, 22(4), 433-454.

Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and instruction in reading comprehension strategies on main idea identification. Journal of Educational Psychology, 83(1), 8-16.

Prior Knowledge

Au, K. (1980). Participation structures in a reading lesson with Hawaiian children. Anthropology and Education Quarterly, 11, 91-115.

Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. (1977). Intrusion of a thematic idea in children's comprehension and retention of stories. Child Development, 48, 1454-1466.

Dewitz, P., Carr, E. M., & Patberg, J. P. (1986). Effects of inference training on comprehension and comprehension monitoring. Reading Research Quarterly, 22, 109-119.

Dole, J. A., Valencia, S. W., Greer, E. A., & Wardrop, J. L. (1991). Effects of two types of prereading instruction on the comprehension of narrative and expository text. Reading Research Quarterly, 26(2), 142-159.

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. Journal of Educational Psychology, 75(6), 821-829.

Manzo, A. V. (1979). The reQuest procedure. Journal of Reading, 2, 123-126.

Linden, M., & Wittrock, M. C. (1981). The teaching of reading comprehension according to the model of generative learning. Reading Research Quarterly, 17(1), 44-57.

Neuman, S. B. (1988). Enhancing children's comprehension through previewing. National Reading Conference Yearbook, 37, 219-224.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 2, 117-175.

Prince, A. T., & Mancus, D. S. (1987). Enriching comprehension: A scheme altered basal reading lesson. Reading Research and Instruction, 27(1), 45-54.

Roberts, T. A. (1988). Development of pre-instruction versus previous experience: Effects on factual and inferential comprehension. Reading Psychology, 9(2), 141-157.

Spires, H. A., Gallini, J., & Riggsbee, J. (1992). Effects of schema-based and text structure-based cues on expository prose comprehension in fourth graders. Journal of Experimental Education, 60(4), 307-320.

Tharp, R. G. (1982). The effective instruction of comprehension: Results and description of the Kamehameha Early Education Program. Reading Research Quarterly, 17(4), 503-527.

Wood, E. G., Winne, P., & Pressley, M. (1988). Elaborative interrogation, imagery, and provided precise elaborations as facilitators of children's learning of arbitrary prose. Paper presented at the American Educational Research Association, New Orleans.

Psycholinguistic

Baumann, J. F. (1986). Teaching third-grade students to comprehend anaphoric relationships: The application of a direct instruction model. Reading Research Quarterly, 21(1), 70-90.

Question Generation (Reviewed by Rosenshine, Meister, & Chapman, 1996)

Signal Word Prompts

Brady, P. I. (1990). Improving the reading comprehension of middle school students through reciprocal teaching and semantic mapping strategies. Unpublished doctoral dissertation, University of Alaska.

Cohen, R. (1983). Students generate questions as an aid to reading comprehension. Reading Teacher, 36, 770-775.

Davey, B., & McBride, M. (1986). Effects of question-generation on reading comprehension. Journal of Educational Psychology, 22, 2-7.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. Elementary School Journal, 90(5), 469-484.

MacGregor, S. K. (1988). Use of self-questioning with a computer-mediated text system and measures of reading performance. Journal of Reading Behavior, 20(2), 131-148.

Palinscar, A. S. (1987). Collaborating for collaborative learning of text comprehension. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 2, 117-175.

Taylor, B. M., & Frye, B. J. (1992). Comprehension strategy instruction in the intermediate grades. Reading Research and Instruction, 32(1), 39-48.

Williamson, R. A. (1989). The effect of reciprocal teaching on student performance gains in third grade basal reading instruction. Unpublished doctoral dissertation, Texas A&M University.

Generic Questions or Question Stems Prompts

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. Contemporary Educational Psychology, 14, 366-381.

King, A. (1990). Improving lecture comprehension: Effects of a metacognitive strategy. Applied Educational Psychology, 29, 331-346.

King, A. (1992). Comparison of self questioning, summarizing, and note taking-review as strategies for learning from lectures. American Educational Research Journal, 29, 303-325.

Main Idea Prompts

Blaha, B. A. (1979). The effects of answering self-generated questions on reading. Unpublished doctoral dissertation, Boston University.

Dreher, M. J., & Gambrell, L. B. (1985). Teaching children to use a self-questioning strategy for studying expository text. Reading Improvement, 22, 2-7.

Lonberger, R. (1988). The effects of training in a self-generated learning strategy on the prose processing abilities of fourth and sixth graders. Unpublished doctoral dissertation, State University of New York at Buffalo.

Ritchie, P. (1985). Graduate research: reviews and commentary: The effects of instruction in main idea and question generation. Reading-Canada-Lecture, 3(2), 139-146.

Wong, Y. L., & Jones, W. (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questioning training. Learning Disability Quarterly, 5, 228-239.

Question Type Prompts

Dermody, M. (1988). Metacognitive strategies for development of reading comprehension for younger children. Paper presented at the American Association of Colleges for Teacher Education, New Orleans.

Labercane, G., & Battle, J. (1987). Cognitive processing strategies, self-esteem, and reading comprehension of learning disabled students. Journal of Special Education, 11, 167-185.

Smith, N. J. (1977). The effects of training teachers to teach students at different reading ability levels to formulate three types of questions on reading comprehension and question generation ability. Unpublished doctoral dissertation, University of Georgia.

Story Grammar Prompts

Nolte, R. Y., & Singer, H. (1985). Active comprehension: Teaching a process of reading comprehension and its effects on reading achievement. Reading Teacher, 39(1), 24-31.

Short, E. J., & Ryan, E. B. (1984). Metacognitive differences between skilled and less skilled readers: remediating deficits through story grammar and attribution training. Journal of Educational Psychology, 76(2), 225-235.

No Prompts

Helfeldt, J. P., & Lalik, R. (1976). Reciprocal student-teacher questioning. Reading Teacher, 33, 283-287.

Manzo, A. V. (1969). Improving reading comprehension through reciprocal teaching. Unpublished doctoral dissertation, Syracuse University.

Simpson, P. S. (1989). The effects of direct training in active comprehension on reading achievement, self-concepts, and reading attitudes of at-risk sixth grade students. Unpublished doctoral dissertation, Texas Technological University.

Other Question Generation Studies (Not Reviewed by Rosenshine et al., 1996)

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. Journal of Educational Psychology, 75(6), 821-829.

Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly, 17(2), 166-186.

Question Answering

Anderson, R., & Biddle, W. (1975). On asking people questions about what they are reading. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 9, pp. 90-132). New York: Academic Press.

Ezell, H. K., et al. (1992). Use of peer-assisted procedures to teach QAR reading comprehension strategies to third-grade children. Education and Treatment of Children, 15(3), 205-227.

Fischer, J. A. (1973). Effects of cue synthesis procedure and post questions on the retention of prose material. Dissertation Abstracts International, 34, 615.

Garner, R., Hare, V. C., Alexander, P. A., Haynes, J., & Winograd, P. (1984). Inducing use of a text lookback strategy among unsuccessful readers. American Educational Research Journal, 21, 789-798.

Garner, R., Macready, G. B., & Wagoner, S. (1984). Readers' acquisition of the components of the text-lookback strategy. Journal of Educational Psychology, 76, 300-309.

Griffey, Q. L., Jr., et al. (1988). The effects of self-questioning and story structure training on the reading comprehension of poor readers. Learning Disabilities Research, 4(1), 45-51.

Levin, J. R., & Pressley, M. (1981). Improving children's prose comprehension: Selected strategies that seem to succeed. In C. M. Santa & B. L. Hayes (Eds.), Children's prose comprehension: Research and practice (pp. 44-71). Newark, DE: International Reading Association.

Pressley, M., & Forrest-Pressley, D. (1985). Questions and children's cognitive processing. In A. C. G. B. Black (Ed.), The psychology of questions (pp. 277-296). Hillsdale, NJ: Erlbaum.

Raphael, T. E., & McKinney, J. (1983). An examination of fifth- and eighth-grade children's question-answering behavior: An instructional study in metacognition. Journal of Reading Behavior, 15(3), 67-86.

Raphael, T. E., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. American Educational Research Journal, 22, 217-235.

Raphael, T. E., & Wonnacott, C. A. (1985). Heightening fourth-grade students' sensitivity to sources of information for answering comprehension questions. Reading Research Quarterly, 20(3), 282-296.

Richmond, M. G. (1976). The relationship of the uniqueness of prose passages to the effect of question placement and question relevance on the acquisition and retention of information. In G. H. Mc Ninch (Ed.), Reflections and investigations on reading. Twenty-fifth Yearbook of the National Reading Conference (pp. 268-278). Clemson, SC: National Reading Conference.

Rowls, M. D. (1976). The facilitative and interactive effects of adjunct questions on retention of eight graders across three prose passages: Dissertation in prose learning. Journal of Educational Psychology, 68, 205-209.

Serenty, M. L., & Dean, R. S. (1986). Interspersed post passage questions and reading comprehension achievement. Journal of Educational Psychology, 78(3), 228-229.

Sheldon, S. A. (1984). Comparison of two teaching methods for reading comprehension. Journal of Research in Reading, 7(1), 41-52.

Watts, G. H. (1973). The "arousal" effect of adjunct questions on recall from prose materials. Australian Journal of Psychology, 25, 81-87.

Wixson, K. K. (1983). Questions about a text: What you ask about is what children learn. Reading Teacher, 37(3), 287-293.

Story Structure

Baumann, J. F., & Bergeron, B. S. (1993). Story map instruction using children's literature: Effects on first graders' comprehension of central narrative elements. Journal of Reading Behavior, 25(4), 407-437.

Buss, R. R., Ratliff, J. L., & Irion, J. C. (1985). Effects of instruction on the use of story structure in comprehension of narrative discourse. National Reading Conference Yearbook, 34, 55-58.

Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. Journal of Reading Behavior, 15(2), 1-17.

Gordon, C. J., & Rennie, B. J. (1987). Restructuring content schemata: An intervention study. Reading Research and Instruction, 26(3), 162-188.

Greenewald, M. J., & Rossing, R. L. (1986). Short-term and long-term effects of story grammar and self-monitoring training on children's story comprehension. National Reading Conference Yearbook, 35, 210-213.

Griffey, Q. L., Jr., et al. (1988). The effects of self-questioning and story structure training on the reading comprehension of poor readers. Learning Disabilities Research, 4(1), 45-51.

Idol, L. (1987). Group story mapping: A comprehension strategy for both skilled and unskilled readers. Journal of Learning Disabilities, 20, 196-205.

Idol, L., & Croll, V. J. (1987). Story-mapping training as a means of improving reading comprehension. Learning Disability Quarterly, 10, 214-229.

Nolte, R. Y., & Singer, H. (1985). Active comprehension: Teaching a process of reading comprehension and its effects on reading achievement. Reading Teacher, 39(1), 24-31.

Omanson, R. C., Beck, I. L., Voss, J. F., McKeown, M. G., et al. (1984). The effects of reading lessons on comprehension: A processing description. Cognition and Instruction, 1(1), 45-67.

Reutzel, D. R. (1984). Story mapping: an alternative approach to communication. Reading World, 24(2), 16-25.

Reutzel, D. R. (1985). Story maps improve comprehension. Reading Teacher, 38(4), 400-404.

Reutzel, D. R. (1986). Clozing in on comprehension: The Cloze story map. Reading Teacher, 39(6), 524-528.

Short, E. J., & Ryan, E. B. (1984). Metacognitive differences between skilled and less skilled readers: Remediating deficits through story grammar and attribution training. Journal of Educational Psychology, 76(2), 225-235.

Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. Reading Research Quarterly, 17(2), 166-186.

Spiegel, D. L., & Fitzgerald, J. (1986). Improving reading comprehension through instruction about story parts. Reading Teacher, 39(7), 676-682.

Varnhagen, C. K., & Goldman, S. R. (1986). Improving comprehension: Causal relations instruction for learning handicapped learners. Reading Teacher, 39(9), 896-904.

Summarization

Afflerbach, P., & Walker, B. (1992). Main idea instruction: An analysis of three basal reader series. Reading Research and Instruction, 32(1), 11-28.

Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? Reading Research Quarterly, 22, 331-346.

Baumann, J. F. (1983). Children's ability to comprehend main ideas in content textbooks. Reading World, 22(4), 322-331.

Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. Reading Research Quarterly, 20(1), 93-115.

Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. Journal of Reading Behavior, 16(4), 297-306.

Berkowitz, S. J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. Reading Research Quarterly, 21(2), 161-178.

Brown, A. L., & Day, J. D. (1983). Macro rules for summarizing texts: The development of expertise. Journal of Verbal Learning and Verbal Behavior, 22, 1-14.

Brown, A. L., Day, J. D., & Jones, R. S. (1983). The development of plans for summarizing texts. Child Development, 48, 968-979.

Carnine, D. W., Kameenui, E. J., & Woolfson, N. (1982). Training of textual dimensions related to text-based inferences. Journal of Reading Behavior, 14(3), 335-340.

Doctorow, M., Wittrock, M., & Marcks, C. (1978). Generative processes in reading comprehension. Journal of Educational Psychology, 70, 109-118.

Jenkins, J. R., Heliotis, J., Stein, M. L., & Haynes, M. (1987). Improving reading comprehension by using paragraph restatements. Exceptional Children, 54, 54-59.

Reutzel, D. R., & Cooter, R. B., Jr. (1988). Research implications for improving basal skill instruction. Reading Horizons, 28(3), 208-215.

Reutzel, D. R., & Hollingsworth, P. M. (1988). Highlighting key vocabulary: A generative-reciprocal procedure for teaching selected inference types. Reading Research Quarterly, 23(3), 358-378.

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. Reading Research Quarterly, 21(4), 422-438.

Sjostrom, C. L., & Hare, V. C. (1984). Teaching high school students to identify main ideas in expository text. Journal of Educational Research, 78(2), 114-118.

Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository material. Journal of Educational Psychology, 74(3), 323-340.

Taylor, B. M. (1986). Teaching middle-grade students to read for main ideas. In J. A. Niles & R. V. Lalik (Eds.). Solving Problems in Literacy: Learners, Teachers, and Researchers: Thirty-fifth Yearbook of the National Reading Conference (pp. 99-108). Rochester, NY: The National Reading Conference.

Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository prose. Reading Research Quarterly, 19, 134-136.

Teacher Training

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. Teaching and Teacher Education, 8(4), 391-403.

Duffy, G. G. (1993). Rethinking strategy instruction: Four teachers' development and their low achievers' understandings. Elementary School Journal, 93(3), 231-247.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliff, G., Book, C., Meloth, M. Vavrus, L., Wesselman, R., Rutnma, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. Reading Research Quarterly, 22(3), 347-368.

Franklin, M. R., et al. (1993). Overcoming the reading comprehension barriers of expository texts. Educational Research Quarterly, 16(1), 5-14.

Vocabulary Comprehension Relationship

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15(1), 3-18.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. Reading Research Quarterly, 20(5), 522-535.

Tomesen, M., & Aarnoutse, C. (1998). Effects of an instructional programme for deriving word meanings. Educational Studies, 24(1), 107-128.

COMPREHENSION II

Appendices

Appendix A

A total of 203 studies met the Panel's criteria for inclusion as scientific studies on comprehension instruction. These studies were grouped into 16 different categories, each representing a particular instructional strategy or collection of strategies. In the following pages, each category of studies is summarized. The Panel defines and describes the rationale for each kind of instructional strategy, the procedures used, and how the instruction was assessed by the researchers. The Panel then evaluates the category of instruction, based on reported results.

Comprehension Monitoring (Also Known as Metacognitive Awareness)

"Comprehension monitoring in the act of reading is the noting of one's successes and failures in developing or attaining meaning, usually with reference to an emerging conception of the meaning of the text as a whole, and adjusting one's reading processes according" (Harris & Hodges, 1995, p. 39). A related concept is "metacognitive awareness," which is "knowing when what one is reading makes sense by monitoring and controlling one's own comprehension" (Harris & Hodges, 1995, p. 153).

Comprehension monitoring, first studied by Markman (1978), involves the readers becoming aware of when they understand what they are reading. Instruction of comprehension monitoring involves teaching readers to become aware of when they do understand, to identify where they do not understand, and to use appropriate fix-up strategies to improve comprehension when it is blocked (Taylor et al., 1992). For reading, comprehension monitoring is "thinking about thinking," an awareness by readers of their ongoing comprehension process while reading. Typically, readers do not spontaneously select comprehension strategy awareness. This instruction strategy involves self-listening (monitoring) or listening to others (Elliott-Faust & Pressley, 1986) and thinking that is designed to help the reader or listener identify when there are problems understanding particular content, such as noticing the comprehension blocks. Comprehension monitoring training is intended to provide readers with steps that they can take to resolve reading problems as they arise. Steps may include formulating what the difficulty is, restating what was read, looking back through the text, and looking forward in the text for information that might help to resolve a problem (Bereiter & Bird, 1985).

The Panel found 20 studies on comprehension monitoring. Table 2, on the following page, summarizes the rationale, procedures, and assessment of research studies on the instruction of comprehension monitoring strategies.

Evaluation

Grade Level

In this search, the Panel found 20 studies on comprehension monitoring instruction. The 20 studies are listed in the bibliography under the rubric Comprehension Monitoring. The distribution of grade levels studied in research on comprehension monitoring ranged from grades 2 to 6: grade level 2, n = 3; level 3, n = 6; level 4, n = 8; level 5, n = 5; level 6, n = 6. Hence, the mode was at grade 4.

Texts

Comprehension monitoring has been studied mainly with expository texts that are used in the elementary grades, particularly social studies and science texts. These present problems with novel concepts and vocabulary as well as novel facts and relationships.

Experimenter Tests

Awareness During Reading

The vast majority of studies on comprehension monitoring investigated whether children could learn to become aware of their comprehension difficulties and verbally report them to the teacher. In terms of success, 16 of 16 studies (100%) measured and obtained more

TABLE 2
COMPREHENSION MONITORING INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The goal of comprehension monitoring is to develop awareness by readers of the cognitive processes involved during reading. Readers learn to become aware of whether they are understanding a text and what steps they should take to correct comprehension difficulties.	The teacher demonstrates awareness of difficulties in understanding words, phrases, clauses, or sentences. Students are taught to: 1. Formulate what it is that is causing them difficulty in understanding. 2. Use think-aloud procedures that show the readers and the teacher where and when understanding difficulties occur. 3. Look back in the text to try to solve a problem. 4. Restate or paraphrase a text in terms more familiar to readers. 5. Look forward ("watch" for information) in a text to solve a problem.	Learning of comprehension monitoring itself. Experimenter tests 1. Detection of inconsistencies in logic of an argument or meaning of a passage. 2. Recall. 3. Long-term maintenance of comprehension monitoring. 4. Self-esteem. 5. Creative thinking. Standard comprehension tests.

success in awareness of comprehension during reading (or listening) for the treatment as compared to the control groups. This success occurs at about the same rate across grades 2 through 6.

Detection of Inconsistencies in Text

Asking the reader to detect inconsistencies in the text is one of the primary means that researchers have used to evaluate success of training and its transfer. Although this is difficult to do, even for adults (Markman, 1983), five studies report significant improvement in error detection for comprehension monitoring conditions.

Other Experimenter Measures

Recall, question answering, and course achievement gains were used once, twice, and once, respectively. The recall and question-answering effects were null for 2nd graders, suggesting that this method does not generalize, at least for the youngest readers. However, one study that measured improvement in science course achievement found that 2nd graders benefited from the training.

Standard Comprehension Tests

Seven studies used standardized comprehension tests to assess general transfer effects of learning comprehension monitoring. Of these, five reported significant effects (grades 3 through 6), and two had no significant effects (grades 3 and 4).

Summary Evaluation of Comprehension Monitoring

Children in grades 2 through 6 can be taught to monitor their comprehension, become aware of when and where they are having difficulty, and learn procedures to assist them in overcoming the problem. There is evidence that this training has specific and general transfer benefits. The main transfer is to improved detection of text inconsistencies and memory for the text and on standardized reading comprehension test performance.

Cooperative Learning

> Cooperative learning is defined as any pattern of classroom organization that allows students to work together to achieve their individual goals (Harris & Hodges, 1995, p. 45).

A related approach is called "collaborative learning," which is defined as "learning by working together in small groups, so as to understand new information or to create a common product" (Harris & Hodges, 1995, p. 35).

As indicated above, cooperative learning involves students working together as partners or in small groups on clearly defined tasks. The tasks require the participation of each student. Mixed ability groups may work together. Readers teach each other. The readers are encouraged to break down the content area material from "teacher-talk" to "kid-talk" to facilitate learning (Klinger, Vaughn, & Schumm, 1998).

Cooperative learning instruction has been successfully used to teach reading comprehension strategies in content subject areas and for teaching across the curriculum. Cooperative learning classes lead to improved academic performance, greater motivation toward learning, and increased time on task (Bramlett, 1994). Students of all abilities benefit from cooperative learning. Furthermore, it has been found to be effective for integrating academically and physically handicapped students into regular classrooms (Klinger et al., 1998).

The majority of teaching, reciprocal teaching, and transactional strategy instruction programs have taken place in small groups rather than large classrooms (Klinger et al., 1998). Cooperative learning is a means for teaching a variety of comprehension strategies in small groups.

The Panel found 10 studies on cooperative learning. Table 3 summarizes the rationale, procedures, and assessment of research studies on cooperative learning and strategy instruction.

Evaluation

Grade Level

The grade levels for cooperative learning were evenly distributed at two each over grades 3 to 6.

Experimenter Tests

The reading strategies that were instructed were successfully learned in the ten studies that measured them. Two studies evaluated the success of the instructional arrangement by analyses of the talk of the children. These analyses showed increased focus on intellectual content and what was being read.

Standardized Tests

Three studies found significant improvement in reading comprehension as measured by standardized tests.

Summary Evaluation of Cooperative Learning

Having peers instruct or interact over the use of reading strategies leads to an increase in the learning of the strategies, promotes intellectual discussion, and increases reading comprehension. This procedure saves on teacher time and gives the students more control over their learning and social interaction with peers.

Curriculum Plus Strategies

Curriculum plus strategy instruction integrates strategy skill training across content areas. A curriculum plus strategy instruction provides the students with cognitive strategy instruction in the context of ongoing academic activities, across school subjects, and throughout the school year. In this approach, each strategy may be taught individually, allowing students to practice a strategy to attain skill. Then students learn to apply the strategies as they need them while reading in each subject area. Individual strategies such as question generation and asking, prediction, clarification, and summarization are taught in conjunction with metacognitive support and flexible use of the strategies (Pressley, Gaskins, Wile, Cunicelli, & Sheridan, 1991).

TABLE 3
COOPERATIVE LEARNING INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The aim of cooperative learning is to teach children to read together with a partner. Readers learn to read aloud with a partner and to listen to the partner's reading. Readers are given activities that teach them strategies for effective reading comprehension.	Students are taught and allowed to participate in partner reading, summarization of paragraphs, and turn-taking in making predictions. Oral reading and listening is done by reader and peers.	Experimenter tests Analyses of peer talk during cooperative learning Summarization Prediction
The readers become independent of the teacher and learn to tutor each other. This reduces the amount of time that the teacher spends with a student.	Training is given, and children learn to carry out activities that follow the self or partner reading, including word recognition (decoding), story structure, prediction, and story summary activities related to texts.	Standardized tests

TABLE 4
CURRICULUM PLUS STRATEGY INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The goal of a curriculum strategy is to provide students with multiple opportunities, in an ongoing school context, to become aware of and develop their cognitive processes across school subjects and throughout the school year.	The focus of these studies is the interaction between teachers and students. The idea behind adding strategic teaching is to attain consistency in this interaction despite variation in content.	Experimenter tests - Comprehension - Monitoring Standardized tests Achievement grades
A curriculum strategy provides students with opportunities to adapt and practice various cognitive strategies in different subjects: reading, writing, social studies, science, and mathematics.	In reading instruction, students are given opportunities to identify text structure. In writing instruction, the students are given opportunities to apply structures. In social studies instruction, students attempt a structural analysis of the texts.	
Experiences that integrate listening, speaking, reading, and writing promote growth in reading and written composition.	Cooperation is encouraged among students working in small groups practicing and applying strategies.	
Motivates students who are at potentially high risk for educational failure.		

The Panel found eight studies on curriculum plus strategies instruction. Table 4, on the following page, summarizes the rationale, procedures, and assessment of research studies on curriculum plus strategies instruction.

Evaluation

The Panel found eight studies that investigated the effects of curriculum experimentally. As noted in Table 4, these studies added strategic instruction to the program of instruction, notably comprehension monitoring, which often differed from standard reading instruction that used basal or directed reading.

Grade Level

The grade levels studied were K through 8 for two of the curriculum investigations. These were literary in nature and focused on real literature rather than basal readers. The remainder of grade levels studies were level 2, n = 1; level 3, n = 2; and level 4, n = 1. These studies used curricula that focused on content areas, literary content, and writing as part of literacy instruction.

Experimenter Tests

General comprehension improvement was reported in seven out of eight studies; four studies reported significant gains in standardized tests. Because instruction in strategy comprehension is a part of the curriculum, it is difficult to assess how the strategies and their learning benefited the readers. Our analysis of multiple strategies and transactional instruction below, however, is consistent with the idea that teaching comprehension strategies as part of the content areas or reading curriculum is an effective procedure.

Summary Evaluation Curriculum Plus Strategies

The variation and complexity of curricula across these studies do not permit one to argue for the scientific support of a particular curriculum or for the particular strategies added to the instruction. However, the success of these individual studies indicates that there may be merit in adding comprehension instruction of reading strategies to a given curriculum and evaluating the results scientifically against those of control groups.

Graphic Organizer

A graph is a "diagram or pictorial device that displays relationships" (Harris & Hodges, 1995, p. 101). In teaching readers to use external means of representing the meaning of relationships in a text, teachers instruct students to organize their ideas through the construction of graphs of ideas based upon what they read, hence the term "graphic organizer."

To help readers construct meanings and organize the ideas presented in a text, the use of graphs or the construction of graphs focuses the readers on concepts and their relations to other concepts. Graphic organizers are methods used to teach the reader to use diagrams of the concepts and their relationships. They are particularly appropriate for expository texts used in content areas such as science or social studies, but they have also been applied to stories as "story maps." The external graphic aids (1) help students focus on text structure while reading, (2) provide tools to examine and visually represent textual relationships, and (3) assist in writing well-organized summaries.

The Panel found 11 studies on graphic organizer instruction. Table 5, on the following page, summarizes the rationale, procedures, and assessment of research studies on graphic organizer instruction.

Evaluation

The Panel found 11 studies that used graphic organizers to assist students in framing and identifying the main ideas in social studies and science texts.

Grade Level

The grade level distribution for the use of graphic organizers is level 2, n = 1; level 3, n = 1; level 4, n = 5; level 5, n = 4; level 6, n = 6; level 7, n = 2; level 8, n = 2. Hence, the modal level is grade 6 with the technique becoming more frequent at grade level 4. Graphic organizing is an activity that is taught to readers in the higher elementary and middle school grades, 4 through 8, with the mode occurring at grade 6. This suggests that children who can learn and benefit from this instruction have to have skill in writing and reading.

TABLE 5
GRAPHIC ORGANIZER INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Readers are instructed to make graphic representations of text material. Graphic organizers include semantic maps, expository maps, story maps, story schema, and graphic metaphors. Graphic organizers visually (spatially) represent super-ordinate and more important subordinate ideas of a passage, story, or exposition. Spatial (graphic) metaphors are assumed to facilitate learning and memory of text and the making of well-organized summaries.	Teachers show readers how to create graphic organization of ideas. Teachers may provide graphic metaphors such as making an umbrella for main ideas and putting details below the topic. or Teachers show readers how to construct maps of expository passages by locating the title or main concept in the center of a circle and then writing in the related ideas from a survey of the text for main ideas. or Teachers show readers how to make box diagrams of a story, for example, problem box-action box-results box and filling in the content of the boxes.	Experimenter tests - Summaries - Text recall Standardized tests - Comprehension subtest of - Gates-MacGinitie Reading Test

Experimenter Tests

Seven studies used recall of the text content to evaluate the effect of training on the use of a graphic organizer. Six of the seven report significant benefits to the experimental groups; one reported a null finding. Four studies (three other than those using recall) report significant achievement gains in the content area. Thus, the main effect of graphic organizers is on improving the reader's memory for the content that is read.

Standardized Tests

Two studies reported positive findings on grades 6 through 8 for standardized tests to evaluate transfer from learning to organize content graphically.

Summary Evaluation of Graphic Organizer Instruction

Teaching students to use a systematic, visual graph to organize the ideas that they are reading about develops the ability of the students to remember what they read and may transfer in general to better comprehension and achievement in social studies and science content areas.

Listening Actively

Listening is the "act of understanding speech." A child's "listening comprehension level" is the "highest grade level of material that can be comprehended well when it is read aloud to the student," also known as "auding, the processes of perceiving, recognizing, interpreting, and responding to oral language" (Harris & Hodges, 1995, p. 140 and p. 14, respectively).

Listening to another person read and following what is being read by reading the text is a method used to teach students how to listen while reading. In the 1970s, efforts were made to train listening skills in general. Dickson (1981) summarizes the relevant work on this kind of training.

Active listening by the student can promote reading comprehension. Students have been taught more effective listening by applying Palinscar's and Brown's (1984) reciprocal teaching (see below) strategies to listening (Grant, 1989). For students in a remedial reading class, listening lessons improved their critical listening, critical reading, and general reading comprehension.

The Panel found four studies on listening instruction and comprehension of text. Table 6, on the following page, summarizes the rationale, procedures, and assessment of research studies on listening instruction.

Evaluation

The Panel found four studies that investigated how listening during reading affects comprehension

Grade Level

Listening studies were carried out on students in grade level 1, n = 1; level 4, n = 1; level 5, n = 1; and level 6, n = 1.

Experimenter Tests

Questions answering showing improvement in two studies.

Standardized Tests

Improvement is reported in two studies on standardized tests.

Summary Evaluation of Listening Instruction

Direct instruction on learning to listen to others (teachers or peers) who read while following in the text what is read may benefit students' comprehension in specific and in more general ways.

Mental Imagery

A mental image is "a-perceptual representation or ideational picture of a perceptual experience, remembered or imagined" (Harris & Hodges, 1995, p. 152).

In imagery training, students are instructed to construct visual images to represent a text as they read it. The text is often a short passage or a sentence. Imagery training improves students' memory (Levin & Divine-Hawkins, 1974) and inferential reasoning about written text (Borduin, 1994).

The Panel found seven studies on mental imagery instruction. Table 7, on the following page, summarizes the rationale, procedures, and assessment of research studies on mental imagery instruction.

TABLE 6
LISTENING ACTIVELY INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Instruction is aimed at achieving active listening for meaning by the reader. Emphasis on listening for meaning produces better sentence recall than emphasis on accurate oral reading. Students who take "active listening turns" are assumed to remember more sentences from a lesson than those who follow along. Listening instruction focuses interest in material. Subject interest is a major factor in sentence recall that is more important than readability. Listening instruction supposedly improves critical listening, reading, and general reading comprehension. It increases participation in group discussions and leads to more thoughtful responses to questions.	The teacher guides the students in critical listening instruction. The teacher poses questions for the students to answer while they listen to the teacher read the text.	Experimental tests - Pretest and posttest on reading and listening Standardized tests - Subtests of Sequential Test of Educational Progress

TABLE 7
MENTAL IMAGERY STRATEGY INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Readers are instructed to make an image to represent the text content. Generating an image requires an interpretation of the text as to its referent(s). When the reader can construct an image of what is read, the reader is assumed to have understood the referent of the text. The constructed image serves as a memory representation of the reader's interpretation of the text.	Teachers ask readers to construct an image(s) that represents the content. This is most often done at the sentence level.	Experimenter Tests - Recall - Short-answer questions

Evaluation

The Panel located seven studies that used mental imagery training and examined its effects experimentally.

Grade Level

Imagery has been used in studies at all grade levels higher than the 2nd grade. The distribution of grades studied was grade level 2, n = 1; level 3, n = 2; level 4, n = 2; level 5, n = 1; level 7, n = 1; and level 8, n = 1. Mental imagery instruction while reading sentences appears to be applicable to grades 2 through 8.

Experimenter Tests

The main effect of imagery is to increase memory for the sentence imaged. The main memory tests used were recall (3 studies) and question answering (6 studies). Keyword cues were used as prompts in five of these studies. In addition, detection of inconsistency showed improvement in two studies.

Summary Evaluation of Mental Imagery Instruction

Instructing readers to imagine what they are reading and coding what they imagine with a keyword cue facilitates readers' memory of what they have read.

Mnemonic Instruction

"Mnemonic procedures include devices or techniques that are aimed at improving memory" (Harris & Hodges, 1995, p. 156).

Mnemonic instruction is a procedure that uses external memory aids. It is a procedure that trains students to use a picture or a concept as a proxy for a person, concept, sentence, or passage. Students are taught to generate an interactive image between the proxy (a word or a picture) and the information covered in the text. This procedure increases learned associations between the proxy and other information in text. The method has been used successfully to teach unfamiliar concepts (e.g., biographies of unfamiliar people, information about unfamiliar places). Although both good and poor readers benefit from this procedure, good readers seem to benefit more (Peters & Levin, 1986). A "keyword" can serve as a proxy.

The Panel's search yielded only two studies on mnemonic instruction and comprehension instruction. Both these studies used keyword methods. Table 8, shown on the following page, summarizes the rationale, procedures, and assessment of research of these studies.

Evaluation

The two studies that used keywords as mnemonics were done on 8th graders. Both found improved recall for passages that had keywords.

Summary Evaluation for Mnemonics

Mnemonic methods using keywords as organizers increase memory and recall. The relationship to other measures of comprehension is not known.

Multiple Strategy Instruction

A "strategy" is "in education, a systematic plan, consciously adapted and monitored, to improve one's performance in learning" (Harris & Hodges, 1995, p. 244). Strategies can be taught and reading requires the flexible use of several different kinds of strategies.

Skilled reading involves the coordinated use of several cognitive strategies. Readers can learn and flexibly coordinate these strategies to construct meaning from texts. Several individual strategies are reviewed in this report. In this section, we examine studies that teach readers to use more than one strategy in the context of reading and in interaction with a teacher over the text. Hence, multiple strategy instruction occurs in a dialog between the teacher and the student. Students are taught individual strategies when and where they are appropriate, usually through modeled use by the teacher. Over the course of reading a passage, several strategies may be taught in conjunction with one another. For example, the reader may predict along with clarification of a word's meaning, activation of knowledge about a story schema, and summarization of the main idea, and all with awareness of problems that are encountered during the reading. In multiple strategy instruction, students are taught how to adapt the strategies and use them flexibly, according to their situation (Pressley, 1991). The teacher models and assists in the learning and flexible use of the strategies by the student. Cooperative learning or peer tutoring may be used as a part of multiple-strategies instruction.

Table 8
Mnemonic Instruction

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The reader is taught to use a keyword to substitute (serve as a proxy) for a person or some aspect of text (person, concept, sentence, passage). The keyword is associated with an interactive image of the referent of a sentence or paragraph. This method is useful when the reader is trying to learn information about totally unfamiliar concepts (e.g., people or countries). The method is assumed to increase the reader's memory through association of the keyword element and other information in the text.	Teacher instructs students to form a keyword substitute for some aspect of prose (person, concept, place, situation, sentence), for example, "tailor" for "Taylor". Pictures are used to help students understand the text. The picture is organized around the keyword.	Experimenter tests - Recall

One variant of multiple-strategy instruction is called "reciprocal teaching." The teacher first models (demonstrates through personal use) and then explains what a strategy is and when to use it (Palinscar & Brown, 1984; Lysynchuk et al., 1990). At first, the teacher guides the reader in applying and practicing strategies while reading a passage. Modeling includes not only examples but the teacher "thinking aloud" to demonstrate the coordinated use of strategies. Gradually, the student begins to practice and implement each strategy independently. In explicit transactional approaches that use multiple strategies, the teacher will explain a strategy before modeling it in a passage (Rosenshine & Meister, 1994).

The Panel found 38 studies on multiple-strategies instruction. Of these, 27 studies were on "reciprocal teaching." The definitions, rationales, procedures, and assessments for "reciprocal teaching" are described in Table 9, on the following page. The 11 studies on other treatments of multiple strategies are summarized in Table 12.

Evaluation of Reciprocal Teaching

Meta-analysis

In "reciprocal teaching," the teacher models by showing how she or he would try to understand the text, using two or more combinations of four strategies: question generation, summarization, clarification, and prediction of what might occur. Rosenshine and Meister (1994) conducted a meta-analysis on 16 reciprocal training studies. Rosenshine and Meister used the criteria of selection that was adopted by us: a study had to be an experimental study with controls and use random assignment or matching of conditions. The grade levels studied were 1 through 8, distributed as level 1, n = 1; level 2, n = 1; level 3, n = 4; level 4, n = 6; level 5, n = 3; level 6, n = 4; level 7, n = 4; and level 8, n = 1. The modal grade for reciprocal teaching was grade 4, but high numbers occur for grades 3 through 7 in these studies (4 on average). Reciprocal teaching using multiple strategies presumes basic reading (decoding) skills, even on those two or more grades below level.

The kinds of strategies included varied from one to four components of summarization, question generation, clarifying, and predicting. Question generation was most frequent (nine studies), followed by summarizing (six studies).

The effect sizes (Rosenshine & Meister, 1994, Table 5, page 194) for experimenter tests (10) studies averaged 0.88; for standardized tests (9 studies), the average effect size was 0.32. These values were about the same for high- and low-quality studies (0.88 and 0.86, respectively, for experimenter tests; 0.31 and 0.36, respectively, for standardized tests). The low-quality studies showed the same effect (0.87) for experimenter tests but a small negative effect (-0.12) for standardized tests. Excluding the low-quality studies, the effect size for standardized tests was raised to 0.36 (seven studies).

Effect size varied as a function of reader ability. Table 11 summarizes these data.

In Table 10, it can be seen that the magnitude of the effect size for experimenter tests was larger for below-average or poor readers. Despite greater efficacy of specific training, scores of standardized tests declined as did the ability of the reader. These data suggest that good readers benefit and generalize what they learn as strategies more than do poor or below-average readers.

Rosenshine and Meister (1994) tested for the significance of effect sizes and examined their results as a function of grade level, excluding below-average readers. These data are summarized in Table 12. Their results show that reciprocal teaching of strategies is not significant for grade 3, is mixed for grades 4, 5, and 6, and is significant for grades 7 and 8. Thus, as measured by significant effect sizes, the older readers benefit most from reciprocal teaching.

Reciprocal Teaching Studies Not Reviewed by Rosenshine & Meister, 1994

The Panel located 11 studies on reciprocal teaching that were not covered in the meta-analysis of Rosenshine and Meister (1994). These studies covered grade levels from 1 to 6 (level 1, n = 1; level 2, n = 1; level 3, n = 3; level 4, n = 3; level 5, n = 3; and level 6, n = 1). These studies tended to use more strategies (seven had combinations of summarization, question generation, clarification, and prediction) and added, in one case each, either monitoring or collaborative learning. Four studies reported improvement on experimenter tests, and three reported significant improvement on standardized tests. These data are consistent with those of Rosenshine and Meister (1994).

TABLE 9
MULTIPLE STRATEGIES: RECIPROCAL TEACHING

TABLE 9
MULTIPLE STRATEGIES: RECIPROCAL TEACHING

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Multiple-strategies instruction is designed to take place in the context of a dialog between the teacher and the students--each of whom reads text passages. In some cases, the teacher also explains a strategy.	The teacher guides the reader in applying and practicing strategies while reading a passage. The teacher models each strategy in the context of reading a passage. The student then applies the strategy to his or her own reading of a passage. There are four main strategies (varies from two to four): 1. Generation of questions during reading 2. Summarization of main ideas of the passage 3. Clarification of word meanings or confusing text 4. Prediction of what might occur later in the text. Optional additions include question answering, making inferences or drawing conclusions, listening, monitoring, thinking aloud,and elaborating.	Experimenter tests - Learning and use of strategies is assessed by analyses of: · Recall · Generating · Answering questions · Summarizing (main idea) · Predicting (what will happen in new passage) Content area achievement Standardized tests

TABLE 10
EFFECT SIZE AS FUNCTION OF READER ABILITY

TYPE OF STUDENT	TYPE OF TEST (number of studies)	
	Standard	Experimenter
All	0.32 (4)	0.85 (5)
Good-Poor	0.19 (2)	0.88 (3)
Below Average	0.08 (4)	1.15 (2)

TABLE 11
EFFECT SIZE SIGNIFICANCE AND GRADE LEVEL

STUDENTS	EFFECT OF GRADE LEVEL		
	Significant	Mixed	Not Significant
Good-poor/All			
3			X
3			X
4		X	
4 & 6	X		
4 & 7		X	
5 & 6		X	
6, 7, 8	X		
7	X		

Table 12
Multiple Strategies: Other Treatment Combinations

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The instruction takes place primarily through the student practicing a given strategy, with feedback from the teacher. The teacher may initially model the strategy.	There are several skills that are practiced here. Packages of skills vary in number from 2 to 5: Self study of the passage. Oral reading Rereading Retelling Review Summarization of main ideas Generation of questions Testing hypotheses Deriving word meaning from morphemes Word recognition training Vocabulary instruction Drawing conclusions Filling in blanks in the passage (Cloze procedure) Monitoring of comprehension Story structure Collaborative learning with partner, including listening to partner reading. Debating or arguing with the author of the text or with the teacher or partner Classification of words, phrases, and sentences	Same as reciprocal teaching

Summary of Reciprocal Teaching of Multiple Strategies

There is strong empirical evidence that the instruction of more than one strategy in a natural context leads to the acquisition and use of these reading strategies and transfers to standard comprehension tests.

Evaluation

Grades

The 12 studies involved readers from grades 2 through 11. The grades were distributed: level 2 = 1, level 3 = 2, level 4 = 6, level 5 = 1, level 6 = 2, and levels 7 through 11, 1 each. Thus, the modal grade is grade 4. Again, basic decoding skill is assumed in teaching reading strategies.

Strategy Instruction

The strategies taught varied across these studies. Six out of the twelve taught summarizing or identifying main ideas. Three had question answering or generation. Monitoring was trained in two studies. Others used cooperative reading, recall, retelling, hypothesis testing, story structure, and psycholinguistic training (word, phrase, and sentence classification, morphological analysis).

Experimenter Tests

Seven studies report specific learning of the strategies taught; two studies report mixed results; and two studies report negative findings. The mixed results and negative findings occurred over grades 4 through 6.

Standardized Tests

No data on standardized tests were reported.

Summary of Other Multiple Strategy Treatment Studies

One or more strategies taught in the context of an interaction facilitates comprehension as evidenced by memory, summarizing, and identifying main ideas.

Overall Summary of Instruction of Multiple Strategies

Taken together, the evidence supports the use of combinations of reading strategies in natural learning situations. These findings build on the empirical validation of strategies alone and attest to their use in the classroom context.

Prior Knowledge

By prior knowledge, the Panel means knowledge that stems from previous experience. This knowledge is a key component of schema theories of reading comprehension (Anderson & Pearson, 1984). Schema theory holds that comprehension depends upon the integration of new knowledge with a network of prior knowledge. Harris and Hodges (1995) offer that within a schema theory, reading is an active process of meaning construction in which the reader connects old knowledge with the new information that is encountered in the text.

To read with understanding, the reader has to have a considerable amount of knowledge. In learning about a content area subject, children acquire knowledge that they can use to understand a text on that content area. In effect, children need prior experience and acquired knowledge to be able to read (Athey, 1983). A reader must activate what he or she knows to use it during reading to comprehend a text. Without activation of what is known that is pertinent to the text, relevant knowledge may not be available during reading, and comprehension may fail; this is analogous to listening to someone speak an unknown foreign language. Teachers can develop relevant knowledge through instruction in content areas prior to reading. One method of reading about other people, in fiction or social studies, asks students to think of their own experiences and how their lives compare with the life situation of someone that is described in a text. This procedure activates relevant prior knowledge and recalls experience that aids understanding (e.g., a trip to the dentist).

A body of work related to prior knowledge activation is called "elaboration interrogation." This procedure encourages students to ask themselves why facts in a text make sense; prior knowledge is stimulated by this

TABLE 13
PRIOR KNOWLEDGE ELICITATION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Students possess considerable knowledge of the world that they can use to comprehend what they are being taught and what they read. Prior knowledge affects comprehension by creating expectations about the content, thus directing attention to relevant parts, enabling the reader to infer and elaborate what is being read, to fill in missing or incomplete information in the text, and to use existing mental structures to construct memory representations that facilitate later use, recall, and reconstruction of text.	Teachers encourage children to compare their lives with situations in the text, either prior or during the reading. Teachers ask students to make predictions about content based on their prior knowledge, often in response to pre-reading questions about the text. Teachers have students practice answering inferential, postreading questions by drawing on text information and prior knowledge. Teachers ask students to search the text and to use what they know to answer inferential questions about the text. Teachers ask students to monitor adequacy of answers to questions on the text.	Recall Short-answer questions (cued recall)

procedure (Martin & Pressley, 1991). This suggests that question elaboration, generation (see below), and answering (see below) are related in that they all necessarily activate and use prior knowledge.

The Panel found 14 studies on prior knowledge instruction. Table 13, on the previous page, summarizes the rationale, procedures, and assessment of research studies on prior knowledge instruction.

Evaluation

Grade Level

The activation and use of what the reader knows that is relevant to what is being read has been studied experimentally for students in grades 1 through 9. The distribution of these grade levels is level 1, n = 1; level 2, n = 2; level 3, n = 1; level 4, n = 6; level 5, n = 2; level 6, n = 2; and level 9, n = 1.

Methods

Most of the studies activated knowledge prior to reading by asking the students to think about topics relevant to the passage to be read (five studies). The remaining studies varied in how prior knowledge was made available: teaching the relevant knowledge (two studies), pre-reading (one study), predicting based on one's own experience (one study), making associations during reading (one study), and previewing the story or text (two studies). Two studies did not specify their methods in the abstracts.

Experimenter Tests

Memory measures were the favored method of assessing comprehension. Recall was used in nine studies, question answering was used in three studies, and achievement in content area was used in two studies. All reported significant effects of prior knowledge on these assessments except for one grade 4 study that previewed the text (Spires, 1992).

Summary Evaluation of Prior Knowledge

The activation of relevant world knowledge helps children understand and remember what they read. The activation of prior knowledge occurs naturally in contexts in which subject content is taught by the teacher, and readers then read text that relates to what has been learned. Prior knowledge activation occurs with several strategies, notably question elaboration, generation, and answering.

Psycholinguistic Instruction

Psycholinguistics is "the interdisciplinary field of psychology and linguistics in which language behavior is examined. Psycholinguistics includes such areas of inquiry as language acquisition, conversational analysis, and the sequencing of themes and topics in discourse" (Harris & Hodges, 1995, p. 197).

The Panel found only one study that trained readers on a psycholinguistic skill, for example, understanding the referents of pronouns. This kind of instruction helps young and developing readers recognize "words that stand for other words" in "anaphoric" relationships, that is, personal pronouns or repeated nouns such as when the word "it" refers to a preceding noun, noun phrase, or clause (Baumann, 1986). Baumann's study on teaching 3rd graders anaphoric reference found that the experimental treatment group increased in accuracy in identifying referents. No transfer or standardized tests were used.

Table 14, on the following page, summarizes the rationale, procedures, and assessment of research studies on psycholinguistic instruction.

Evaluation

Grades

The one study involved readers from grade 3.

Summary Evaluation of Psycholinguistic Training

Children may need some instruction in reading contexts to aid them in establishing who is being referred to by personal pronouns. Instruction apparently does work. The lack of studies in this area suggests that much more training on syntactic and semantic relationships could be developed and researched for its effectiveness.

Table 14 Psycholinguistic Strategy

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Readers need to learn that words that stand for or refer to other words, e.g., "she" stands for a female referent introduced earlier in the text.	Teachers model or show readers how to identify the antecedents of pronouns and to answer questions based on identified antecedents.	Experimenter Tests Students answer pronoun-specific questions after reading expository or narrative texts.
This strategy is used to communicate the use of a word or phrase that stands for a preceding word or phase, like a pronoun.	Readers learn to identify noun substitutes, verb substitutes, and clause substitutes.	Students write down the antecedents for underlined anaphoric terms in expository text.
Readers come to understand the semantic relationship between a pronoun and the word or phrase to which it refers.		

Question Answering

When queried by teachers, themselves, or others, young readers experience difficulty in answering questions well. Question-answering instruction is intended to aid students in learning to answer questions while reading and thus learn more from a text. Students can also learn procedures for answering questions or what to do when they cannot answer a question. If students can develop these strategies, their learning from text is facilitated when the answers are available in the text.

There were 17 studies on question answering instruction. Table 15, on the following page, summarizes the rationale, procedures, and assessment of research studies on question-answering instruction.

Evaluation

Grade Level

Question answering begins with students in grade 3 and has been studied up to grade 8. The distribution of reported grade levels is level 3, n = 2; level 4, n = 3; level 5, n = 3; level 6, n = 1; and level 8, n = 1. The preponderance of studies, then, has been on grades 3 through 5.

Experimenter Tests

Improvement in performance by treatment vs. control groups is reported on question answering (nine studies), looking back in text (three studies), question generation (one study), and recall (one study).

Standardized Tests

There are no reports on the use of standardized tests in abstracts of the question answering studies surveyed.

Summary of Evaluation of Question Answering

Instruction of question answering leads to an improvement in answering questions after reading passages and in strategies of finding answers. This improvement occurs in grades 3 through 8. The effects of this method, however, are small.

Question Generation

The goal of reading strategy instruction, in general, is to teach readers to become independent, active readers who use strategies that enhance their comprehension. One strategy that achieves this goal is question generation in which the reader learns to pose and answer questions about what is being read. Without

TABLE 15
QUESTION ANSWERING

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Question-answering strategy instruction assists students learning from a text. A question focuses the student on particular content and can facilitate reasoning (e.g., answering why or how). In content questions, the information available in the text determines, in great part, the student's ability to answer the questions. Teaching students to look back in the text when they cannot answer a question facilitates their learning. Students can learn to discriminate questions that can be answered based on the text vs. those that are based on their own knowledge and require the generation of inferences or conclusions. Questions after the reading of a passage can lead to reprocessing of relevant text after the reader fails to answer the question.	Teachers ask students questions during or after reading passages of text. Teachers ask students to look back to find answers to questions that they cannot answer after one reading. Teachers ask students to analyze questions with respect to whether the question is tapping literal information covered in the text, information that can be inferred by combining information in the text, or information in the reader's prior knowledge base. Questions often come at the end of science and social studies or in workbooks to accompany texts. These may be used in question answering.	Experimenter Tests - Recall - Short answer questions - Look back in text to answer question

training, young readers are not likely to question themselves. Nor are they likely to use questions spontaneously to make inferences. The assumption of question generation instruction is that readers will learn to engage text by making queries that lead to the construction of better memory representations. The goal is to teach students to make these self-questions while reading. If one asks why, how, when, where, which, and who kinds of questions, it is possible to integrate segments of text, to thereby improve reading comprehension and memory for what is read, and to gain a deeper understanding of the text. Question generation should also increase the reader's awareness of whether the text is being understood. When the teacher is present, the reader's creation of questions may signal success or failure in comprehension and prompt the teacher or the reader to attempt to compensate for comprehension failure. Finally, question generation has been studied in isolation or as a multiple-strategy instruction program such as reciprocal teaching.

In the Panel's search, it located a recent literature review on question generation by Rosenshine, Meister, and Chapman (1996). Rosenshine and his colleagues conducted a meta-analysis of 30 studies that instructed students how to generate questions during reading, either as a single strategy or in combination with other reading strategies. Of these, 11 studies used the "reciprocal teaching" method, and question generation was part of a set of two or more strategies that were taught. These studies were described in Table 10 above. Nineteen additional studies reviewed by Rosenshine et al. (1996) investigated instruction of question generation alone or in combination with strategies not taught by reciprocal teaching methods.

The Panel found 27 studies on question generation instruction. Table 16, on the following page, summarizes the rationale, procedures, and assessment of research studies on question generation instruction.

Evaluation

The main evaluation of question generation is based on the meta-analysis of Rosenshine, Meister, and Chapman (1996) who employed the same criteria as Rosenshine and Meister (1994) for selection of studies.

Grade Level

The study of question generation instruction begins with grade 3 and has been carried out up to grade 9. The distribution of grade levels in this study of this kind of instruction is level 3, n = 3; level 4, n = 6; level 5, n = 4; level 6, n = 9; level 7, n = 4; level 8, n = 3; level 9, n = 2. The modal level is grade 6.

Experimenter Tests

The respective effect sizes for multiple choice (n = 6), short-answer (n = 14), and summary (n = 3) measures were 0.95, 0.85, and 0.85.

Standardized Tests

The median effect size for 13 studies that used standardized comprehension tests was 0.36. The median effect sizes for standardized vs. experimenter tests are reported in Table 17 (following Table 16), broken down by reciprocal teaching and other treatments. The magnitude of the median effect sizes in Table 17 is approximately the same as that found for reciprocal teaching of multiple strategies. There is an overlap of studies here so that the similarity is likely a result of common studies. It is of interest that although there is a positive effect size for standardized tests, only 3 out of 13 are statistically significant. Experimenter tests fare better here because 16 out of 19 are statistically significant. Thus the effects of instruction of question generation are specific to learning the particular strategy and may not generalize to standardized tests.

Summary Evaluation of Question Generation

There is strong empirical and scientific evidence that instruction of question generation during reading benefits reading comprehension in terms of memory and answering questions based on text as well as integrating and identifying main ideas through summarization. There is mixed evidence that general reading comprehension improved on standardized comprehension tests. Question generation may be best used as a part of a multiple-strategy instruction program.

Story Structure

A story is "an imaginative tale shorter than a novel but with a plot, characters, and setting, as a short story." A "story map" is "a time line showing the ordered sequence of events in a text" or "a semantic map showing the meaning of relationships between events or concepts in the text, regardless of their order." (Harris & Hodges, 1995, pp. 243-244). Story structure refers to the finding in discourse analysis that the content of stories is systematically organized into episodes and that the plot of a story is a set of episodes. Knowledge of episodic content (setting, initiating events, internal reactions, goals, attempts, and outcomes) helps the reader understand the who, what, where, when, and why of stories as well as what happened and what was done.

Story structure instruction is a method by which the teacher teaches the reader knowledge and procedures for identifying the content of the story and the way it is organized into a plot structure. In addition to learning the episodic content, the reader can learn to infer causal and other relationships between sentences that contain the content. This learning gives the reader knowledge and procedures for deeper understanding of stories and allows the reader to construct more coherent memory representations of what occurred in the story.

Table 16
Question Generation Instruction

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The goal of question generation is to teach readers to become independent, active, self-questioners. The assumption is that readers will learn more and construct better memory representation when self-questions are asked while reading. Integrative questions that capture large units of meaning should improve reading comprehension and memory of text by making readers more active while reading. Question generation is often a part of a multiple-strategy program such as reciprocal teaching. Question generation should increase students' awareness of whether they are comprehending text.	Teachers ask children to generate questions during the reading of a passage. The questions should integrate information across different parts of the passage. Teachers ask children to evaluate their questions about whether the questions covered important material, were integrative and could be answered based on what is in the text. Teachers provide feedback on the quality of the questions asked or assist students in answering the questions generated. Teachers teach the students to evaluate whether their questions covered important information, whether the questions were integrative, and whether they themselves could answer the questions.	Experimenter tests - Quality of questions generated - Question answering Standardized comprehension tests

TABLE 17 MEDIAN EFFECT SIZES FOR QUESTION GENERATION
(DATA FROM ROSENSHINE, MEISTER, & CHAPMAN, 1996, APPENDIX D)

RECIPROCAL TEACHING N = 11 STUDIES	STANDARDIZED TESTS	EXPERIMENTER TESTS
Median Effect Size	0.34 (n = 6)	0.88 (n = 7)
Number Significant	0 out of 6	7 of 7
OTHER TREATMENTS		
Median Effect Size	0.35 (n = 7)	0.82 (n = 12)
Number Significant	3 of 7	9 of 12

TABLE 18
STORY STRUCTURE INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
Instruction is aimed at teaching the student how stories and their plots are organized into episodes. Readers know a great deal about the content and structure of stories as a genre. However, training in how stories and their plots are organized into episodes can aid a reader in understanding the who, what, where, when, and why of narratives. Stories often entail problems that are faced by people, and they provide a context in which students can learn about problemsolving by experiencing the lives of others. Asking and answering the questions of who, what, when, where, and why, as well as learning about problems and their solutions, are useful procedures that are trans-situational and apply to stories as well as to real life. Knowing the structure of the story and its time, place, characters, problems, goals, solutions, and resolution facilitates comprehension and memory for stories. Stories constitute the bulk of the texts used in elementary school reading.	Teachers teach students to ask and answer five questions: 1. Who is the main character? 2. Where and when did the story occur? 3. What did the main characters do? 4. How did the story end? 5. How did the main character feel? Students learn to identify the main characters of the story, where and when the story took place, what the main characters did, how the story ended, and how the main characters felt. Students learn to construct a story map recording the setting, problem, goal, action, and outcome over time. Students construct a story map while reading stories. Some mapping procedures require recording the setting, problem, goal, action, and outcome information.	Experimenter tests - Retell the story (recall) - Short-answer questions

The Panel found 17 studies on story structure instruction. Table 18, which follows Table 17, summarizes the rationale, procedures, and assessment of research studies on story structure instruction.

Evaluation

Grade Level

Research on story structure instruction begins in grade 3, n = 2, but increases in grade 4, n = 8 (four studies on poor readers). This trend continues into grade 5, n = 7 (on poor readers), and grade 6, n = 2 (on poor readers).

Experimenter Tests

The main kinds of tests used to evaluate experimental training on story structure are recall (n = 10 successes and 1 failure in grade 5 among normal readers), question answering on the stories (n = 8 successes, and 1 failure in grade 5 among normal readers), and identifying the elements of a story structure (n = 5 successes and 2 failures: 1 in grade 3 and 1 in grade 5, both with normal readers). All studies on poor readers report improvement on experimenter tests.

Standardized Tests

Three studies report the use of standardized tests following training in story structure. There were two successes and one failure (grade 5, normal readers).

Summary Evaluation of Story Structure Instruction

Instruction in the content and organization of stories improves comprehension of stories as measured by the ability of the reader to answer questions and recall what was read. This improvement is more marked for less able readers. More able readers may already know what a story is about and therefore do not benefit as much from the training. However, this kind of instruction aids both kinds of readers.

Summarization

A summary is "a brief statement that contains the essential ideas of a longer passage or selection" (Harris & Hodges, 1995, p. 247). To be able to create a summary of what one has just read, one must discern the most central and important ideas in the text. One also must be able to generalize from examples or from things that are repeated. In addition, one has to ignore irrelevant details.

The assumption in teaching students how to summarize what they read is that most students do not summarize well. The central aim of most summarization instruction is to teach the reader how to identify the main or central ideas of a paragraph or a series of paragraphs.

Summarization training is effective. It can be transferred to situations requiring general reading comprehension, and it leads to improved written summaries. Summarization training can make students more aware of the way a text is structured and how ideas are related. If asked to summarize, students have to pay closer attention to the text while they read. They also learn to spend more time on reading and trying to understand what they read. In some instances, training increases the quality of students' note taking and recall of major information (Rinehart, 1986).

The Panel found 18 studies on summarization instruction. Table 19, on the following page, summarizes the rationale, procedures, and assessment of research studies on summarization instruction.

Evaluation

Grade Level

Summarization instruction studies are rare below grades 5 and 6. Of those reporting information on grades studied, we found one level 3 and one level 4. There were four and nine studies on grades 5 and 6, respectively. There was one study at the high school level. Summarization often presupposes writing as well as reading skill. This may be one reason for its use for upper elementary school grades.

Experimenter Tests

The majority of the studies reported improvement of the quality of summaries (n = 11). Other studies reported improved recall of what was summarized (n = 7) and improved question answering (n = 4). No negative findings were reported.

Standardized Tests

Standardized tests were rarely used. Only two studies reported using them on 6th graders; one succeeded and the other failed in increasing comprehension.

TABLE 19
SUMMARIZATION INSTRUCTION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The aim of summarization instruction is to teach the reader to identify the main or central ideas of a paragraph or a series of paragraphs. To do so, the reader needs to use prior knowledge of the content of the text as well as knowledge of grammar. Furthermore, the reader has to make inferences that go across sentences and beyond the text. The reader must learn to generalize. Integrating text through main ideas leads to a more organized, succinct, and coherent memory representation of what was read.	Readers are taught to summarize paragraphs by rule application, mainly to delete trivial and redundant information; to use superordinates; and to identify or generate a main idea. The reader is taught through example and feedback to apply any of five rules: 1. Deletion of trivia 2. Deletion of redundancy 3. Superordination, which replaces a list of exemplars with a superordinate term 4. Selection of a topic sentence to serve as a scaffold of the summary 5. Invention of a topic sentence for a paragraph where one was not explicitly stated. Readers gain experience in summarizing single- or multiple-paragraph passages. With multiple paragraphs, readers first summarize individual paragraphs and then construct a summary of summaries or a spatial organization of the paragraph summaries.	Recall of expository or narrative text Question answering with open or multiple-choice answers

Summary Evaluation of Summarization

The instruction of summarization succeeds in that readers improve the quality of their summaries of text, mainly in identifying the main idea but also in leaving out detail, including ideas related to the main idea, generalizing, and removing redundancy. This result indicates that summarizing is a good method of integrating ideas and generalizing from the text information. Furthermore, instruction in summarization improves memory of what is read, both in free recall and in answering questions. This strategy of instruction is used as part of reciprocal teaching and other treatments that teach multiple strategies. It is an important component.

Teacher Preparation for Text Comprehension Instruction

Teachers have to learn how to teach reading comprehension strategies and procedures. Teachers can do this by becoming more aware of, and being prepared on, the procedures and processes of good comprehension of text. Teachers need to learn how to interact with students during the reading of a text to teach them reading comprehension strategies at the right time and right place. The goal of teacher preparation for text comprehension instruction is to provide teachers with opportunities to learn about the cognitive processes that occur in reading, how to instruct in comprehension strategies that can be utilized by the reader, how to teach strategies through demonstration and other techniques, how to explain them, how to allow the student to learn and use them in the context of reading a text, and how to use individual strategies in conjunction with several other reading comprehension strategies.

Teacher preparation on strategy instruction is recent and rare. When teachers receive and implement training on strategy instruction, reading comprehension improves. The idea of the teacher as a modeler of thinking strategies and as a coach facilitating them is new. As a result, few teachers have received practical preparation in the teaching of cognitive strategy instruction (Anderson & Roit, 1993; Duffy, 1993).

Four studies were found on teacher preparation instruction. Table 20, on the following page, summarizes the rationale, procedures, and assessment of these research studies. The next section of this report conveys a more detailed analysis of preparation of teachers in strategies, focusing on recent, successful programs that occur in natural reading contexts involving transactions among the reader, teacher, and text.

Evaluation

Grade Levels

Teachers were prepared to teach students multiple strategies for text comprehension from grades 2 through 11. The distribution is fairly uniform over this range of grades. Of interest is the fact that all the studies, save one (Franklin, 1993), were carried out on "poor readers," "disabled students," or "low achievers."

Experimenter Tests

With respect to the teachers' learning and faithfulness to the treatment, all six studies claim success. With respect to student benefits from the teachers who were prepared in instructing multiple reading strategies, two studies report improvement in the subject matter of the instruction.

Standardized Tests

Two studies report success in improving performance on standardized comprehension tests.

Summary Evaluation of Teacher Preparation to Teach Text Comprehension

This is a very important area for study. To implement the teaching of reading strategies in naturalistic classroom environments, it is important to know how and whether teachers can be effectively prepared in instructional procedures. Furthermore, it is important to learn about time and other costs that are associated with such instruction. Finally, it is important to determine whether students as well as teachers learn and benefit from the teacher preparation. This small set of studies indicates that teachers can learn to implement comprehension strategy instruction in the classroom under natural teaching circumstances. It also suggests that students benefit from such instruction by prepared teachers. There is a need to carry out additional preparation studies of this kind with a wider range of readers. Normal readers, as well as others who are less skilled in reading, could benefit from implementation of the teaching of multiple reading comprehension strategies, not only in reading instruction but in content areas as well.

Table 20
Teacher Preparation on Comprehension Instruction

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The aim of teacher preparation is to instruct teachers in teaching reading comprehension strategies in the classroom context and in natural interaction with students.	Teachers undergo preparation in multiple strategies and explanation of strategies. Teachers are instructed in strategic reading techniques and a collaborative transactional approach to reading informational texts. Teachers are prepared to make decisions and explain mental processing associated with reading skills as strategies. Self-evaluative workshops are often used for learning and feedback. Teachers also learn from the use of transcripts of lessons, videos, and post-lesson interviews.	Experimenter tests Fidelity to treatment by teachers: - Do teachers learn and teach the strategies in which they were trained? - Videotape pre- and posttests - Reading sessions Comprehension by students: - Do students learn and practice the strategies taught? - Do students show gains in reading comprehension? Awareness of lesson content Achievement in content learning Standardized reading tests

TABLE 21
VOCABULARY INSTRUCTION AND RELATION TO COMPREHENSION

DEFINITION AND RATIONALE OF INSTRUCTION	PROCEDURES TAUGHT OR PRACTICED	ASSESSMENT
The aim of vocabulary instruction is to use instruction and reciprocal teaching methods to teach strategies for discovering the meanings of unfamiliar words.	Teacher models being a "word detective," looking for contextual clues to find word meaning, a synonym, or an antonym by analyzing words and word parts and by looking at surrounding text description for clues to meaning.	Experimenter Tests - Word meanings - Cloze tests Standardized tests
Intensive vocabulary instruction is designed to promote word knowledge that will enhance text comprehension.	Teachers elaborate on word meanings and use them in diverse contexts, adding activities to extend use of learned words beyond the classroom.	
	The learning tasks provide definitions, knowledge, fluent access to word meanings, context interpretation, and story comprehension.	
	Students encounter words multiple times (16 to 20), highlight and use vocabulary terms to generate inferences, complete sentence stems, generate contents or situations appropriate to target words, and fill in words that are missing in a Cloze procedure.	

Vocabulary Instruction and Relation to Comprehension

Vocabulary knowledge is correlated with reading comprehension (see the Comprehension I report). The rationale and procedures for teaching vocabulary are found in Beck, Perfetti, and McKeown (1982).

The instruction of vocabulary and assessment of learning vocabulary with respect to comprehension can show whether this correlation is, in fact, causal. Although the first section of the subcommittee report shows that vocabulary can be acquired through instruction, few of those studies examined whether successful instruction of vocabulary leads to increased comprehension. Four studies were found on vocabulary-comprehension instruction. Table 21, which follows Table 20, summarizes the rationale, procedures, and assessment of research studies on vocabulary and its relation to comprehension instruction.

Evaluation

The Panel found two studies by McKeown (1983, 1984) on teaching vocabulary that also assessed students on comprehension. These 4th grade students were tested on word meanings, Cloze procedures, and story comprehension. The author reports success in learning of the words and use of word meanings and in increased story comprehension. In addition, there is a study by Tomeson and Aarnouste (1998), who applied reciprocal teaching methods to teach vocabulary to 4th grade students. Students learned to derive word meanings from text, but transfer to more general reading comprehension as assessed by a Dutch standardized test was not successful.

Summary Evaluation of Vocabulary Instruction and Relation to Comprehension

More experimental studies on the relationship between learning vocabulary and reading comprehension are needed. There is a high correlation between vocabulary knowledge and comprehension. Is there a causal direction between learning vocabulary and improving reading comprehension? Furthermore, vocabulary learning is a part of normal content area learning. Instruction in vocabulary in content areas may lead to better reading and listening comprehension and to improvement in course achievement. This is a promising area of research because it bridges early reading skill development and later comprehension training.

Appendix B

This Appendix summarizes information on three questions:

- What are the claims in the literature about the effectiveness of instruction on comprehension?
- What grades have been studied?
- What are some of the implications for instruction in the classroom?

Table 22, on the following page, provides information on the 16 categories of instruction to answer these questions. For each category, there are sections that describe the effects claimed by the researchers, the grade levels that were studied, and ways in which the method might be taught in a classroom setting.

TABLE 22
RELEVANCE OF INSTRUCTION

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Comprehension Monitoring	Children can be taught to monitor their comprehension and become aware of when and where they are having difficulty during reading. They can learn procedures to assist them in overcoming the problem that they are having withunderstanding what they are reading. This training has specific and general transfer benefits. The main transfer is to improved detection of text inconsistencies and memory for the text and improved performance on standardized reading comprehension tests.	2 to 6	Comprehension monitoring can be taught through teacher modeling of the process and practice by children in doing it during reading. Comprehension monitoring can be taught in natural reading contexts where children read aloud and have difficulty with word recognition or word and sentence meaning. Teachers can be trained on how to teach comprehension modeling either preservice or inservice. They can be taught how to think aloud and to communicate their own understanding processes to the students. The students can learn with feedback to look back or forward in the text and to use the text to find clues as to the meaning of words and sentences. Comprehension monitoring can be taught as a part of a larger program of reading strategies in interaction with the teacher in natural reading or content areas.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Cooperative Learning	When students as peers tutor or instruct one another or interact over the use of reading strategies, the evidence is that they learn reading strategies. They engage in intellectual discussion, and they increase their reading comprehension. This procedure develops independent learning by children and frees the teacher for other activities and students. The students gain more control over their learning and social interaction with peers. The study of cooperative learning in natural reading contexts and as a part of a program of instruction that uses multiple strategies needs to be done. Teacher training studies on how to teach cooperative learning in natural reading contexts need to be done.	3 to 6	Cooperative learning or peer tutoring can be developed in group reading situations where students work together to learn and use reading comprehension strategies. Cooperative learning can be a part of a natural reading program where peers as well as the teacher engage in a transaction over the meaning of a text in a content area or in reading instruction. Teachers can be trained on how to develop cooperative learning, either in experimental investigations, or in preservice or inservice development.

Table 22
Relevance of Instruction (continued)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Curriculum	The variation and complexity of curricula across these studies does not permit one to argue for the scientific support of a particular curriculum nor for the particular strategies added to the instruction Because the kinds of strategies added to a given curriculum works when studied in isolation or as a part of a set of multiple strategies, adding them to an existing reading curriculum or to content area curricula should enhance learning, comprehension, and course achievement.	2 to 4	Teachers can be trained in instruction of a variety of strategies. They can learn to teach these strategies in reading or content area instruction. Teacher preparation studies are needed to assess their fidelity to treatment and the effectiveness of the strategies as part of a curriculum. Fidelity of the students' learning of the strategies needs to be assessed in natural reading or content area instruction The relationships of teacher preparation and student learning of strategies needs to be assessed in terms of general transfer to comprehension tests, but, more importantly, to improved content area achievement.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Graphic Organizer	Teaching students to use external aids and writing to organize their ideas about what they are reading is a proven procedure that enhances comprehension for text. The use of systematic, visual or semantic graphs on the content of a passage benefits the student in terms of better memory for what was read. Furthermore, this preparation, when done in Social Studies and Science content areas, facilitates memory and content area achievement. Teaching teachers to use graphic organizers has not been studied. The use of graphic organizers as a part of a reading instruction program has not been studied.	2 to 8	Teachers could be trained to teach students how to graphically represent ideas and relations for either narrative or expository text while reading in either a natural reading or content area instructional context. Studies on teacher preparation and student learning, fidelity to treatment, and general comprehension effects of this procedure in natural contexts and as a part of a package of strategies needs to be studied. Teacher preparation on the use of this strategy could be done preservice or inservice.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Listening	Instruction on learning to listen to others (teachers or peers) while reading may benefit readers' comprehension in specific and in more general ways. The number of studies on listening is small, and listening's effectiveness lacks a strong scientific base. Teaching teachers to teach students how to listen to the teacher and to peers who read orally needs to be studied further. It is likely that listening occurs informally as part of reading and content area instruction.	1 to 6	Teachers can be trained to teach students listening skills when the teacher or peers read. The teacher assesses comprehension through questioning. Fidelity to treatment of teachers and students needs to be assessed in studies of the effectiveness of instruction on listening during reading. Instruction on listening during reading could be added to instruction of a package of reading comprehension strategies in the teaching of reading or content area

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Mental Imagery	Instructing readers to imagine what they are reading and coding what they imagine with a keyword cue facilitates readers' memory what they have read. This method is useful for imagining the referents of individual sentences. This method seems to be limited to memory for particular sentences No studies on preparation of teachers or students on the use of imagery in reading or content areas have been done.	2 to 8	The use of imagery is an easy strategy to teach. Teachers could be trained to use it appropriately at sentences during the reading of text in natural reading or content areas. This method would actively engage the reader to use mental processes that lead to good recall. Furthermore, it could be used during oral reading and listening because imagery is easier when listening than when reading. This strategy could be added to a repertoire of strategies.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Mnemonic	This method is similar to graphic organizers (Pressley et al., 1989). The use by students or teachers of keywords or concepts to organize main ideas and relationships or to generalize from instances can lead to better specific memory. The use of an external referent such as a picture has limited utility.	8 Its use in other grades has not been studied. It is similar to graphic organizers that have proven use in grades 2 to 8.	Teachers could be taught to use words as concepts or classes to help students organize ideas that are subordinate or related to main ideas. This teaching could be part of an instruction program in reading or in a content area.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Multiple Strategies	There is very strong empirical, scientific evidence that the instruction of more than one strategies in a natural context leads to the acquisition and use of these reading strategies and transfers to standard comprehension tests. Preparation of teachers in the use of multiple strategies in interactive instruction has been successful (see Teacher Preparation below).	3 to 8	Teachers can be trained in the use of multiple strategy instruction in natural reading or content areas. Current programs of transactional research are promising examples of this. Fidelity to treatment by both teachers and students is desired and should be studied. Studies need to be done on when, where, and how to implement strategy instruction in natural instructional contexts. Teachers could be trained on multiple reading strategy instruction in-service or pre-service. The instruction of multiple reading strategies should not be restricted to poor reader.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Prior Knowledge	The activation of relevant world knowledge helps children understand and remember what they read. The activation of prior knowledge occurs naturally in contexts where subject content is taught by the teacher and readers then read text that relates to what has been learned. It is not clear that this procedure has to be explicitly taught, especially in content areas.	2 to 6	Teachers teach content areas in a variety of ways which provide the kind of knowledge that readers can later activate to understand the current text. Prior knowledge studies indicate that prior learning or learning that precedes reading enhances comprehension of what is read. In this sense, reading about a subject after learning about it in other ways would be a part of a program of instruction in a content area. Research on how learning content prior to reading about it and its benefits needs to be studied.

Table 22
Relevance of Instruction (continued)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Psycholinguistic	Children may need some instruction in reading contexts to aid them in establishing who is being referred to by personal pronouns. Instruction apparently does work here. The lack of studies here suggests that much more training on syntactic and semantic relationships could be developed and researched for its effectiveness.	4 Its use in other grades has not been studied. Instruction on how to use knowledge of syntax, semantics, text properties, and text genre has not been done.	Teachers might benefit from preparation in linguistic and discourse analyses and how to teach children how to deal with complexity of sentences and genres. This has been successfully done with stories as a genre (see Story Structure below). Children need more experience in early exposure to expository (non-narrative texts) so that they can learn properties and strategies of coping with this kind of text. This is best done by earlier introduction to texts on Science and Social Studies. Teachers could teach children about understanding complexities of sentences and different genres by their adoption earlier in the reading and content area curricula. The teaching of understanding of these kinds of texts would involve the use of modeling of as well as sue of procedures for teaching other reading comprehension strategies.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Question Generation	There is strong empirical and scientific evidence that instruction of question generation during reading benefits reading comprehension in terms of memory and answering questions based upon text as well as integrating and identifying main ideas through summarization. There is mixed evidence that general reading comprehension is improved on standardized comprehension tests. Question Generation may be best used as a part of a multiple strategy instruction program. Question Generation enables the student to be actively involved in reading and to be motivated by his own queries rather than those of the teacher in question answering.	3 to 9	Question generation should be part of a program of instruction of reading comprehension strategies in a natural reading or content area context. Teachers can be taught to ask readers to generate questions and to provide feedback in these contexts. Students can learn to generate and find answers to their own questions. Fidelity to treatment by teachers and students needs to be assessed. The relation of successful learning needs to be related to content area achievement as well as standardized tests.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Question Answering	Instruction of Question Answering leads to an improvement in answering questions after reading passages and in strategies of finding answers.	3 to 8	Question asking by teachers and question answering by students is a part of natural reading and content area instruction. It should be explicitly taught to teachers with the addition that they give feedback on answers and elaborate the feedback in the context of the text or content area being taught. Question asking and feedback on the content of the answer should be made a part of programs that give instruction of multiple reading comprehension strategies. Teacher and student preparation on question answering, feedback, and ways to find information that answer questions should be studied in natural instructional contexts on reading and content areas. Teachers could be trained inservice or preservice.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Story Structure	The instruction of the content and organization of stories improves comprehension as measured by the ability of the reader to answer questions and recall what was read. This improvement is more marked for less able readers. More able readers may already know what a story is about and therefore do not benefit as much from the preparation. However, this kind of instruction aids both kinds of readers.	3 to 6	Teachers can be prepared to teach story structure through the use of questions and graphic organizers (story maps). They should not teach story grammar categories per se but rather should focus on the characters, the settings, what happened, how characters felt, what they thought, what they wanted to do, what they did, and how things turned out. When the reading material is narrative, question answering and generation strategies can be used by teachers to draw out the content and organization of stories crucial to the student building a representation of the episodic structure and causal relationships. The use of questions to learn story structure can be a part of a program of instruction of comprehension strategies in natural reading or content areas.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Summarization	The instruction of summarization succeeds in that readers improve on the quality of their summaries of text, mainly identifying the main idea but also in leaving out detail, including ideas related to the main idea, generalizing, and removing redundancy. Summarizing is a good method of integrating ideas and generalizing from the text information. Instruction of summarization improves memory for what is read, both in terms of free recall and answering questions. This strategy instruction has been used as a part of reciprocal teaching and other treatments that teach multiple strategies. It is an important component.	3 to 6	Rules and procedures for the summarization of single and multiple passages can be taught to teachers either inservice or preservice. It is an important strategy for integration and generalization of information found in a text. It is an integral part of multiple strategy instruction and has been widely implemented and studied.

TABLE 22
RELEVANCE OF INSTRUCTION (CONTINUED)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Teacher Preparation	This is a very important area for study. In order to implement the teaching of reading strategies in naturalistic classroom environments, it is important to know how and whether teachers can be effectively prepared in the instructional procedures. Further, it is important to learn about the time and other costs that are associated with such instruction. Finally, it is important to determine whether the students as well as the teachers learn and benefit from the teacher preparation. The small set of studies on teacher preparation indicate that teachers can learn to implement multiple comprehension strategy instruction in the classroom under natural teaching circumstances. The research also suggests that students benefit from such instruction by prepared teachers.	2 to 11 Mostly on poor readers. There is a need for studies on normal and above average readers.	There is a need to carry out further preparation studies of this kind and on a wider range of readers in natural reading and content area instruction. These preparation studies should focus on the implementation of the teaching of several kinds of reading comprehension strategies that have been proven singly or multiply in scientific studies. This implementation should be done in natural occurring contexts, especially in content areas. Normal readers as well as others who are less skilled in reading could benefit from implementation of the teaching of multiple reading comprehension strategies, not only in reading instruction, but in content areas as well. Fidelity to treatment by teachers and students needs to be assessed. The relation of successful learning and teaching by teachers and of successful learning and use of strategies to content area achievement needs to be assessed rather than transfer to general reading comprehension tests.

Table 22
Relevance of Instruction (continued)

TYPE OF INSTRUCTION	HOW EFFECTIVE?	GRADE LEVELS STUDIED	HOW TAUGHT?
Vocabulary-Comprehension	Three studies on instruction report increased word meaning and improvements on experimenter tests of story comprehension or standardized comprehension tests.	4 (see initial section of report on Vocabulary Instruction for a wider range of grades)	Teachers can be prepared to teach word meanings and strategies to create them while reading. Students can learn vocabulary through instruction of word meanings in the context of reading instruction or content area instruction. Basic and classroom research on vocabulary instruction, its effectiveness and its relationship to reading comprehension is needed.

CHAPTER 4

Comprehension

Part III

Teacher Preparation and Comprehension Strategies Instruction

COMPREHENSION III

Teacher Preparation and Comprehension Strategies Instruction

Introduction

The purpose of this subreport is to review what is currently the most promising research direction in the area: the preparation of teachers to deliver comprehension instruction. If further research in this direction is pursued, it is likely to lead to progress in our understanding of reading comprehension instruction, and it will also contribute to the general area of teacher preparation.

Background

Reading comprehension strategy instruction has been a major research topic for more than 20 years. The idea behind this approach to instruction is that reading comprehension can be improved by teaching students to use specific cognitive strategies or to reason strategically when they encounter barriers to comprehension as they read. The earliest work in this area used a "direct instruction" model, in which teachers taught a specific strategy or set of strategies to students. The goal of such training was, as it always is, the achievement of competent and self-regulated reading.

At first, investigators focused on teaching students one strategy at a time. A wide variety of strategies was studied, including imagery, question-generating, prediction, and a host of others. In this approach, teachers usually modeled the cognitive strategies in question, often by "thinking aloud" as they read to demonstrate what proficient readers do. The approach also involved guided practice in which students were led to the point where they were able to perform independently, via a gradual reduction of scaffolding. This type of instruction was effective in helping students acquire the strategy, and usually there was some evidence that the use of the strategy improved performance on reading comprehension tasks. In later studies, several strategies were taught in combination, and these studies showed similar effects. Recommendations to use particular combinations of strategies in actual teaching situations became common.

There are many additional questions that might be asked of the existing literature on single- and multiple-strategy instruction, and many loose ends that could be tied up. For example, few of the existing studies address issues of long-term maintenance of strategy use. Effects of strategy instruction on real reading tasks (e.g., reading connected text) are not well delineated, and there is little evidence on the issues that one typically pursues after the initial experimental forays into a topic, for example, the optimal age for training, how long training should last, and so on.

However, the pursuit of these sorts of detail questions within the context of the work already done might not be the most productive focus for future research because implementation of the direct instruction approach to cognitive strategy instruction in the context of the actual classroom has proved problematic. For one thing, it is often difficult to communicate what is meant by "teaching strategies and not skills." Several papers have been written whose purpose is to explicate exactly *how* teachers are taught to become teachers of comprehension strategies, and it appears that no small part of the challenge of training teachers comes from the difficulty of describing what is required of them. In addition, acquiring and practicing individual strategies in isolation and then attempting to provide transfer opportunities during the reading of connected text makes for rigid and awkward instruction.

Proficient reading involves much more than utilizing individual strategies; it involves a constant, ongoing adaptation of many cognitive processes. To help develop these processes in their students, teachers must be skillful in their instruction. Indeed, successful teachers of reading comprehension must respond flexibly and opportunistically to students' needs for instructive feedback as they read. To be able to do this, teachers themselves must have a firm grasp not only of the strategies that they are teaching the children but also of *instructional* strategies that they can employ to achieve their goal. Many teachers find this type of teaching a challenge, most likely because they have not been prepared to do such teaching. Thus, although the literature on cognitive strategy instruction for reading comprehension has yielded valuable information, it has not provided a satisfactory model for effective instruction as it occurs in the classroom.

The area within comprehension strategy instruction that currently seems to have the most potential for moving the field along is teacher preparation. In this report, the NRP discusses four studies in which teachers are trained to teach strategies and in which the focus is the effectiveness of that training on students' reading. Four studies is not a large number; but it is not surprising that only a few relevant studies have been done. Interest in the topic is rather new, and preparing teachers to deliver effective strategy instruction is a lengthy process.

Methodology

Database

The NRP searched the ERIC and PsycINFO databases to locate relevant studies conducted since 1980. The search terms used were "comprehension," "strategy," and "instruction." There were 453 articles. In addition, the Panel searched using the terms "direct explanation" and "teacher explanation"; this added 182 nonoverlapping items. Recent research reviews were also examined: Lysynchuk, Pressley, d'Ailly, Smith, and Cake (1989), Pressley (1998), Rosenshine and Meister (1994), and Rosenshine, Meister, and Chapman (1996); these reviews did not identify any relevant studies that the searches had not revealed.

Analysis

To be included, a study had to be

- Focused on the preparation of teachers for conducting reading comprehension strategy instruction.
- Published in a scientific journal.
- Empirical.
- Experimental using random assignment or quasi-experimental with initial matching on the basis of reading comprehension scores.
- Comprehensive in reporting the complete set of results of the study. (Ancillary articles that focused on specific aspects of the same database were not included but are listed in the References.)

Four studies met these criteria. A detailed outline of each of the selected studies, organized to permit comparison across studies, is presented in Appendix A.

Our Panel subcommittee reviewed the research in reading comprehension instruction broadly and also selected certain specific topics for a deeper focus, e.g., vocabulary and teacher preparation for teaching reading comprehension strategies. It should be noted that there are other relevant aspects of comprehension instruction, for example, instruction in listening comprehension and in writing, that were not addressed. In addition, the Panel subcommittee did not focus on special populations such as children whose first language is not English and children with learning disabilities. It did not review the research evidence concerning special populations and thus cannot say that its conclusions are relevant to them.

Consistency With the Methodology of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. A formal meta-analysis was not possible because of the small number of studies identified. However, comprehensive summaries according to NRP guidelines for each of the four studies appears in Appendix B.

Results

The results of the selected studies suggest that, in fact, good teacher preparation can result in the delivery of instruction that leads to improvements in students' reading comprehension. However, the variations among the four studies to be discussed here raise questions about what the best approach to teaching teachers to do strategy instruction might be.

There have been two major approaches to comprehension strategy instruction: Direct Explanation (DE) and Transactional Strategy Instruction (TSI). Two studies that represent each approach are described.

Direct Explanation

The Direct Explanation approach was designed to improve on the standard direct instruction approach to strategy instruction used in most of the early studies, in which students are simply taught to use one or several strategies as described above. Arguing that direct instruction was insufficient because it did not attempt to provide students with an understanding of the reasoning and mental processes involved in reading strategically, Duffy, Roehler, and colleagues (1986) developed the DE approach. In this approach, teachers do not teach individual strategies but focus instead on helping students to (1) view reading as a problem-solving task that necessitates the use of strategic thinking and (2) learn to think strategically about solving reading comprehension problems. The focus in DE is on developing teachers' ability to explain the reasoning and mental processes involved in successful reading comprehension in an explicit manner, hence the use of the term "direct explanation." The implementation of DE requires specific and intensive teacher training on how to teach the traditional reading comprehension skills found in basal readers *as strategies*, for example, to teach students the skill of how to find the main idea by casting it as a problem-solving task and reasoning about it strategically.

> Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. Reading Research Quarterly, 21(3), 237-252.

The first study done by Duffy and Roehler's research team investigated whether training teachers to be explicit in their teaching of reading strategies would be effective in increasing the explicitness of their verbal explanations and whether this explicitness would be related to students' meta-cognitive awareness of strategies and to their achievement. Twenty-two teachers were randomly assigned to either the treatment or the control condition. Treatment teachers were trained to use an explanation model that was designed to help them explain reading strategies explicitly to their 5th grade students in low-level reading groups. After an initial training session, the treatment teachers received 10 hours of additional training spaced throughout the school year. During these training sessions, the explanation model was described, and teachers designed lessons according to the model. Their teaching was observed and discussed on four occasions. Control teachers participated in a workshop on classroom management at the start of the study and received no further training. The results of this study indicated that students of teachers who received training in the use of the explanation model had significantly greater awareness of (1) what strategies were taught, (2) why they are important, and (3) how they are used than did students of the comparison teachers.

The Duffy et al. (1986) study thus demonstrated the effectiveness of training teachers, and it showed that explicit explanations by teachers can lead to greater general awareness among students of reading strategies. However, the question of the extent to which students were able to apply these strategies and ways of thinking to their actual reading practice, that is, whether the use of such methods leads to significant improvements in reading comprehension performance, was not answered positively. The treatment and the comparison classrooms did not differ on the posttest administration of the comprehension subtest of the Gates-MacGinitie Test.

> Duffy and colleagues (1986) did find, however, that students of the treatment teachers spent significantly more time answering the items on the comprehension test than did the other students. This suggested to them that perhaps these students were being more thoughtful and strategic in their reading.

There is little point in adapting new teaching methods if they are not shown to be effective in improving actual performance. Thus, the 1986 study by Duffy and colleagues cannot be considered conclusive about the value of training teachers to provide explicit explanations about how to read strategically. However, the results were promising enough to persuade the same research team to undertake another study, similar to this one in many respects, but incorporating a more elaborate program of teacher preparation.

> Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. Reading Research Quarterly, 23(3), 347-368.

In a 1987 study (Duffy et al.), as in the Duffy and colleagues 1986 study, there was random assignment of teachers to condition. Treatment teachers were shown how to provide explicit explanations, in this case to 3rd grade low-level reading students. In addition, the teachers were trained to analyze the skills prescribed in their basal reading texts and to recast these skills as problem-solving strategies. In essence, the emphasis in this study was on the effects of training teachers to provide students with explicit descriptive information about the types of reasoning and mental processes that are used strategically by skilled readers, as opposed to simple prescriptions of how to perform the basal text skills. Included in the 12 hours of training were one-on-one coaching, collaborative sharing among the teachers, observation of lessons and feedback, and videotaped model lessons. Comparison teachers were trained in classroom management and used management principles throughout the study.

The effectiveness of this approach was measured in terms of both student awareness and student achievement. Student awareness of strategic reasoning was assessed in interviews conducted both immediately following lessons and at the end of the yearlong treatment. As in the Duffy and colleagues (1986) study, the results indicated that, compared with students of untrained teachers, the students of trained teachers had higher levels of awareness of specific reading strategies, as well as a greater awareness of the need to be strategic when reading.

The fact that students have high awareness of the reasoning associated with strategic reading does not necessarily mean that they are proficient in using such strategies and better in reading comprehension. Duffy et al. (1987) designed an achievement measure to assess both students' ability to use the basal skills they had been taught and the degree to which their responses reflected the reasoning associated with using skills as strategies. Results indicated that there was no difference between students of treatment and control teachers in the ability to use the skills. However, the students of treatment teachers were found to have a greater ability to reason strategically when reading. Results on a task involving paragraph reading also indicated that students of treatment teachers (1) reported that they used such reasoning when actually reading connected text, and (2) described the reasoning employed when using the strategies. In contrast, students of control teachers were unable to do so.

The 1987 study also used standardized measures to assess students' reading performance. The comprehension and word skills subtests of the Stanford Achievement Test (SAT) were used. Overall, students of the treatment teachers outperformed the others on the posttest. This difference was significant for the word skills subtest but was not significant for the comprehension subtest. A second standardized test, the Michigan Educational Assessment Program (MEAP), was administered as a delayed posttest, to assess whether the overall advantage of students of treatment teachers persisted over time. It was found that even 5 months after the instruction ended, students of the trained teachers had significantly higher reading scores than students of the control teachers.

The results of these two investigations of the DE approach to comprehension strategy instruction suggest that although this approach is clearly useful for increasing student awareness of the need to think strategically while reading, the effects on actual reading comprehension ability are less clearcut. As noted above, both of the Duffy and colleagues studies produced only mixed results on the standardized measures of reading performance. It should be noted, however, that the 1987 study reported that many of their lessons were oriented toward acquisition of word-level processes and not to what are usually considered comprehension processes.

Transactional Strategy Instruction

The TSI approach includes the same key elements as the DE approach, but it takes a somewhat different view of the role of the teacher in strategy instruction. Whereas emphasis in DE is on teachers' ability to provide explicit explanations, the TSI approach focuses not only on that but also on the ability of teachers to facilitate discussions in which students (1) collaborate to form joint interpretations of text and (2) explicitly discuss the mental processes and cognitive strategies that are involved in comprehension. In other words, although TSI teachers do provide their students with explicit explanations of strategic mental processes used in reading, the emphasis is on the interactive exchange among learners in the classroom, hence use of the term "transactional."

In both DE and TSI, teachers explain specific strategies to students and model the reasoning associated with their use. Both approaches include the use of systematic practice of new skills, as well as scaffolded support, in which teachers gradually withdraw the amount of assistance they offer to students. Perhaps the most salient distinction to be made between DE and TSI is the manner in which the different emphases of the two approaches (explanation vs. discussion) result in differences in the level of collaboration among students that takes place in each approach. In the DE approach, strategy instruction is primarily conducted by the teacher. In contrast, the TSI approach is more collaborative: Although explicit teacher explanation is an important part of this approach, TSI is designed for learning to occur primarily through the interactive transactions among the students during classroom discussion.

> Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. Teaching and Teacher Education, 8(4) 391-403.

Anderson (1992) worked with experienced teachers of severely reading-disabled adolescent students. The students ranged from grades 6 through 11, but three-quarters of them had incoming reading levels of grade 3 or below. The teachers were randomly assigned to a treatment or control condition. The nine treatment teachers received three 3-hour sessions of training in the use of the TSI approach, held at intervals during the period during which the actual reading intervention with the students was going on. Special features of Anderson's teacher preparation included (1) the involvement of the teachers as coresearchers who were part of the development of the project and (2) the availability of a previously trained peer coach for each teacher throughout the project.

In their training, the teachers were given a list of changes, or "shifts," that need to be made in most classrooms for more active reading to be fostered. This list of 20 teacher shifts and 12 student shifts first described ways in which teachers and students typically behave during remedial reading instruction and then described contrasting behaviors that characterize or promote active reading. The teachers were also given a set of principles for fostering active reading through reading instruction with specific teaching techniques for each principle. Each treatment teacher was also assigned a previously trained teacher for peer support. There were seven comparison teachers, who received no training.

In the intervention, both teacher groups taught reading comprehension for 3 months, using expository texts. The instruction in treatment classrooms emphasized both direct explanation and collaborative discussion. To evaluate the effects of the TSI approach, the phonics, structural analysis, and reading comprehension subtests of the Stanford Diagnostic Reading Test were administered. There was no difference from pretest to posttest in the performance of students of the trained and untrained teachers on the phonics and structural analysis subtests. However, significantly more students of the trained teachers (80%) made gains on the reading comprehension subtest than did students of the other teachers (50%), suggesting that preparation given the teachers was effective in improving reading comprehension performance. The amount of gain was not reported.

> Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. Journal of Educational Psychology, 88(1), 18-37.

Over the past decade, Pressley and associates have developed a transactional strategy instruction program called Students Achieving Independent Learning (SAIL). In SAIL, reading processes are taught as strategies through direct explanation, teacher modeling, coaching, and scaffolded practice. An important feature of the program is its emphasis on collaborative discussion among teacher and students, including extended interpretive discussions of text, with these discussions emphasizing student application of strategies. A goal of the SAIL program is for students to develop more personalized and integrative understanding of text.

A yearlong study by Brown, Pressley, Van Meter, and Schuder (1996) provides evidence of the effectiveness of the TSI approach as exemplified by the SAIL program. In this study, SAIL was contrasted with a more traditional approach to reading instruction. There was no specific teacher preparation within the context of this study; the five SAIL teachers had all been previously trained and had at least 3 years of experience as SAIL teachers. The five comparison teachers had even more years of teaching experience than the SAIL teachers had, but they had no SAIL training. The students in this study were in 2nd grade; all were reading below grade level at the beginning of the study.

The SAIL teachers and comparison teachers were matched on a variety of measures to form five pairs. In each pair of classrooms, data were collected on six low-achieving students from each classroom who were matched on the basis of their reading comprehension scores. Thus, Brown and colleagues (1996) did the careful matching required when doing a quasi-experiment.

Students' strategy awareness was assessed through interviews. Students of SAIL teachers reported more awareness of comprehension and word-level strategies than did students of comparison teachers (operationalized as the number of strategies they claimed to use during reading). In an evaluation of story recall, the SAIL students did better on literal recall of story content and also were more interpretive in their recalls. On a think-aloud task, SAIL students used more strategies on their own than did the other students. Student reading achievement was also assessed, using the comprehension and word skills subtests of the Stanford Achievement Test. Over the course of the study, students of the SAIL teachers showed greater improvement than the students of the other teachers, and at posttest, they significantly outperformed the others on both subtests.

Discussion

Every one of these studies reported significant differences, and although none of them reported effect sizes, they provided enough information so that effect sizes could be calculated for most of the effects. The effect sizes were substantial, suggesting that these initial attempts to provide effective instruction for teachers in reading comprehension strategy training are promising and worth following up.

It is encouraging to see that random assignment is indeed feasible in these real-life classroom situations. This statement is not intended as a criticism of Brown and colleagues' quasi-experiment, which was done carefully and which, in fact, posed a question that could not be tested in a true experiment: What is the effect of a particular model of instruction (TSI) delivered by teachers experienced and committed to it, working in the context of schools also committed to that approach? This is an important question. But most of the relevant research questions do not demand a quasi-experimental design, and therefore a much better choice would be a true experiment. Sometimes researchers argue that school administrators refuse to allow random assignment because it disrupts their schools. Perhaps researchers should make serious and sincere efforts to find schools that *will* cooperate, because they do exist; and researchers should also help the field by making an effort to educate school administrators about random assignment and other important design standards.

These comments should not be taken as implying that it is easy to do classroom-based naturalistic studies of the type discussed here. It is difficult, and the difficulty should not be minimized. Such research cannot be undertaken without substantial funding and adequate institutional support. It also requires collaboration among researchers; school personnel, including both teachers and administrators; and parents, which does not come about quickly—it requires time and effort. And doing this type of research takes commitment and energy. The research team must remain motivated and effective during a lengthy developmental phase and then during the study itself. Moreover, a high-quality study of this type has probably been preceded by descriptive and correlational work. The emphasis on the importance of experimental studies should not be interpreted as negating the valuable contributions of these other research paradigms in preparing to do intervention research.

Of course, any evaluation of these instructional approaches is limited by the fact that these studies cannot easily be compared. They differed in terms of specific purpose, teacher preparation method, intervention, type of student (age, reading level, etc.), control group, and other characteristics. Nevertheless, taken together, the studies do indicate that instructional methods that generate high levels of student involvement and engagement during reading can have positive effects on reading comprehension. The classroom procedures in each of the studies required substantial cognitive activity on the part of the students. Also, these studies demonstrate that providing teachers with instruction that helps them use such methods leads to students' awareness of strategies and use of strategies, which can in turn lead to improved reading comprehension.

These findings beg the question as to what it is, in fact, that makes for effective strategy instruction. Is it the teacher preparation? (If so, how extensive does it have to be? Would the teachers maintain their instructional effectiveness without the supports inherent in an ongoing study?) Is it the use of direct explanation and/or collaborative discussion when teaching students? Is it the particular strategies that are taught, or would a broader repertoire of instructional activities also be effective? Is it a combination of some or all of these possibilities or of other factors not mentioned here? Clearly, more research is warranted on this topic. In light of the findings to date, one can expect that further work in this area will yield valuable knowledge concerning optimal conditions for improvement in reading comprehension.

Thus, the results of the research to date represent significant progress in our understanding of the nature of reading comprehension and of how to teach it. There is much more to learn, of course. What we must remember is that reading comprehension is extremely complex and that teaching reading comprehension is also extremely complex. The work of the researchers discussed here makes this clear. They have not recommended an "instructional package" that can be prescribed for all students. They have not identified a specific set of instructional procedures that teachers can follow routinely. Indeed, they have found that reading comprehension instruction cannot be routinized. What they have shown, and this is an important new direction in which to take our research efforts, is that intensive instruction of teachers can prepare them to teach reading comprehension strategically and that such teaching can lead students to greater awareness of what it means to be a strategic reader and to the goal of improved comprehension.

Implications for Reading Instruction

General guidelines for teachers that derive from the research evidence on comprehension instruction with normal children include the suggestions that teachers help students by explaining fully what it is they are teaching: what to do, why, how, and when; by modeling their own thinking processes; by encouraging students to ask questions and discuss possible answers among themselves; and by keeping students engaged in their reading via providing tasks that demand active involvement.

The current dearth of comprehension instruction research at the primary grade level should not lead to the conclusion that such instruction should be neglected during the important period when children are mastering phonics and word recognition and developing reading fluency.

In evaluating the effectiveness of strategy instruction in the classroom, the primary focus must be not on the students' performance of the strategies themselves. The appropriate assessment is of the students' reading achievement and, in addition, other outcome measures such as how interested students are in reading and how satisfied teachers are with their instructional methods.

Implementation of effective comprehension instruction is not a simple matter; substantial teacher preparation is usually required for teachers to become successful at teaching comprehension.

There is a need for greater emphasis in teacher education on the teaching of reading comprehension. Such instruction should begin at the preservice level, and it should be extensive, especially with respect to teaching teachers how to teach comprehension strategies.

Conclusions From the Research on Comprehension Instruction

1. The most active topic in the research on comprehension instruction over the last few years has been comprehension strategies instruction with normal children.

2. Teaching strategies for reading comprehension in normal children leads to increased awareness and use of the strategies, improved performance on commonly used comprehension measures, and, sometimes, higher scores on-standardized tests of reading.

3. For further progress to be made, research is needed that focuses on ways that strategies can be taught within the natural setting of the classroom and for both normal children and those with reading difficulties. Work of this type is enhanced when cognitive researchers collaborate with researchers knowledgeable about teacher education.

Conclusions From the Research on Teacher Preparation and Comprehension Strategies

1. Teachers can be taught to teach comprehension strategies effectively; after such instruction, their proficiency is greater, and this leads to improved performance on the part of their students on awareness and use of the strategies, to improved performance on commonly used comprehension measures, and, sometimes, higher scores on standardized tests of reading.

2. Teaching comprehension strategies effectively in the natural setting of the classroom involves a level of proficiency and flexibility that often requires substantial and intensive teacher preparation.

Directions for Further Research

Research evidence suggests that further work in the area of comprehension instruction, on the topic of strategy instruction as well as on other topics, will lead to even more progress. Following is a list of issues that deserve further consideration.

1. Our understanding of the complex construct of reading comprehension has been expanded and refined in our recent research, but the construct is still not completely understood. Studies incorporate a large variety of heterogeneous measures derived from tasks ranging from those requiring simple recognition and recall, through making inferences, to using text information in solving problems and performing other complex tasks. There is no "map" of the construct that investigates relationships among the various methods of defining and measuring comprehension and that determines which measures are optimal for evaluating performance in research studies and in assessing student achievement in the school context.

2. Many investigators do not describe fully all important aspects of their studies—the reader, the text and other materials, the task, and the teacher (see Methodology in Chapter 1 of this volume). An excellent discussion of methodological and reporting standards to ensure high-quality studies is available in Lysynchuk, Pressley, d'Ailly, Smith, and Cake (1989).

3. A variety of methodologies, including descriptive and correlational procedures, will contribute to our knowledge, but intervention research requires experimental studies, using wherever possible a true experimental design, that is, random assignment. Quasi-experiments are acceptable when the specific purpose of the study demands such a design but not when done simply for convenience or ease of implementation.

4. The relationship of comprehension to word-level processes and fluency has not been well investigated.

5. It will be important to know the effects of interventions aimed at increasing motivation.

6. Research should extend to students at the secondary level as well as to children with reading difficulties. Study skills instruction traditionally given to normally achieving and above-average students should be compared to the newer cognitive strategy instruction.

7. There is little research at the K to 2nd grade level on teaching reading comprehension. One important topic at this level is the relationship between listening comprehension and reading comprehension.

8. The research base is scanty with respect to the development of effective methods of vocabulary instruction, especially methods that incorporate direct instruction, how these might vary across age and reading levels and abilities, and how vocabulary training can be integrated optimally with other types of comprehension instruction.
9. Research is needed on how writing is related to reading comprehension.
10. It will be important to develop further the use of videotapes, technology in general, and other techniques for teacher preparation.
11. There is little evidence from cost-benefit analyses to determine the amount of gain in student achievement (and other outcome measures) relative to the cost of implementing a reading comprehension instructional program.
12. With respect to comprehension strategy instruction and teacher preparation:

Comprehension Strategy Instruction: Maintenance and Transfer

1. Do teaching comprehension strategies have lasting effects on students?
2. Do the effects generalize to other reading situations, such as content area instruction?
3. Can comprehension instruction be done successfully within the context of content area instruction?

Teacher Preparation

1. How much teacher preparation is required for successful performance?
2. How should teacher preparation be conducted at the preservice and at the inservice levels?
3. Can teachers maintain their proficiency after their own preparation to teach comprehension has been completed?
4. Does the fact that teachers are involved in an ongoing research study make a difference in their performance?

Other Important Concerns

1. Teacher characteristics

How does a teacher's age, amount of teaching experience, type of preservice education, or other characteristics affect success in comprehension instruction? Which components of successful teacher preparation programs are the effective ones? What characteristics of the teacher preparation itself (its focus, its intensity, its timing) affect the success of a teacher preparation program?

2. Reader characteristics

How do a student's age, reading level, learning ability, proficiency in English, or other characteristics affect success in comprehension instruction?

3. Text characteristics

Does the difficulty level of the texts used in instruction make a difference?

Can one expect transfer from one text genre to another (e.g., from narrative to expository text)?

4. Task characteristics

What characteristics of the instruction delivered to the students are the effective ones? The direct explanation? The collaborative discussion? The particular strategies and tasks taught to the students? The amount of instruction? The active involvement on the part of the students? Other factors?

References

Studies Included in Report

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. Teaching and Teacher Education, 8(4) 391-403.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. Journal of Educational Psychology, 88(1), 18-37.

Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J., & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. Reading Research Quarterly, 21(3), 237-252.

Duffy, G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. Reading Research Quarterly, 23(3), 347-368.

Other Relevant Papers

Book, C. et al. (1985). A study of the relationship between teacher explanation and student metacognitive awareness during reading instruction. Communication Education, 34(1), 29-36.

Casazza, M. E. (1993). Using a model of direct instruction to teach summary writing in a college reading class. Journal of Reading, 37(3), 202-208.

Collins, C. (1991). Reading instruction that increases thinking abilities. Journal of Reading, 34(7), 510-516.

Collins Block, C. (1993). Strategy instruction in a literature-based reading program. Elementary School Journal, 94(2), 139-150.

Collins Block, C. (1999). Comprehension: Crafting understanding. In L. Gambrell, L. Morrow, S. Neuman, & M. Pressley (Eds.), Best practices in literacy instruction (pp. 98-118). New York: Guilford Publications.

Dole, J., Duffy, G., Roehler, L., Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. Review of Educational Research, 61(2), 239-264.

Duffy, G. (Ed.). (1990). Reading in the Middle School, 2nd Ed. Newark, DE: International Reading Association.

Duffy, G., Roehler, L., & Mason, J. (Eds.). (1984). Comprehension instruction: Perspectives and suggestions. New York: Longman.

Duffy, G., Roehler, L., Meloth, M., & Vavrus, L. (1986). Conceptualizing instructional explanation. Teaching and Teacher Education, 2(3), 197-214.

Duffy, G., Roehler, L., Meloth, M., Polin, R., Rackliffe, G., Tracy, A., & Vavrus, L. (1987). Developing and evaluating measures associated with strategic reading. Journal of Reading Behavior, 19(3), 223-246.

Duffy, G., Roehler, L., & Rackliffe, G. (1986). How teachers' instructional talk influences students' understanding of lesson content. Elementary School Journal, 87(1), 3-16.

Duffy, G. (1991). What counts in teacher education? Dilemmas in educating empowered teachers. In J. Zutell & S. McCormick (Eds.). Learning factors/teacher factors: Literacy research and instruction. Fortieth Yearbook of the National Reading Conference (pp. 1-18). Chicago: National Reading Conference.

Duffy, G. (1993). Rethinking strategy instruction: Four teachers' development and their low achievers' understandings. Elementary School Journal, 93(3), 231-247.

Duffy, G. (1997). Powerful models or powerful teachers? An argument for teacher-as-entrepreneur. In S. Stahl & D. Hayes (Eds.), Instructional models in reading (pp. 351-363). Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis, E. S. (1993). Integrative strategy instruction: A potential model for teaching content area subjects to adolescents with learning disabilities. Journal of Learning Disabilities, 26 (6), 358-383, 398.

Fairhurst, M. A. (1981). Satisfactory explanations in the primary school. Journal of Philosophy of Education, 15(2), 205-214.

Frager, A., & Hahn, A. (1988). Turning contemporary reading research into instructional practice. Reading Horizons, 28(4) 263-271.

Hahn, A. (1987). Developing a rationale for diagnosing and remediating strategic-text behaviors. Reading Psychology, 8(2), 83-92.

Heller, M. F. (1986). How do you know what you know? Metacognitive modeling in the content areas. Journal of Reading, 29(5) 415-422.

Herrmann, B. A. (1988). Two approaches for helping poor readers become more strategic. Reading Teacher, 42(1), 24-28.

Herrmann, B. A. (1989). Characteristics of explicit and less explicit explanations of mathematical problem solving strategies. Reading Research and Instruction, 28(3), 1-17.

Lysynchuk, L., Pressley, M., d'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. Reading Research Quarterly, 24(4), 458-470.

Palincsar, A.S., David, Y. M., Winn, J. A., & Stevens, D. D. (1991). Examining the context of strategy instruction. RASE: Remedial and Special Education, 12(3), 43-53.

Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. New York: The Guilford Press.

Pressley, M., El-Dinary, P. B., Gaskins, I., Schuder, T., Bergman, J., Almasi, J., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. Elementary School Journal, 92(5), 513-555.

Pressley, M., Johnson, C., Symons, S., McGoldrick, J., & Kurita, J. (1989). Strategies that improve children's memory and comprehension of text. Elementary School Journal, 90(1), 3-32.

Pressley, M., & Wharton-Mcdonald, R. (1997). Skilled comprehension and its development through instruction. School Psychology Review, 26(3), 448-466.

Pressley, M., Wharton-McDonald, R., Rankin, J., Mistretta, J., Yokoi, L., & Ettenberger, S. (1996) The nature of outstanding primary grades literacy instruction. In E. McIntyre & M. Pressley (Eds.), Balanced instruction: Strategies and skills in whole language (pp. 251-276). Norwood, MA: Christopher-Gordon.

Pressley, M., Yokoi, L., Rankin, J., Wharton-McDonald, R., & Menke, D. (1997). A survey of the instructional practices of grade-5 teachers nominated as effective in promoting literacy. Scientific Studies of Reading, 1(2), 145-160.

Roehler, L., & Duffy, G. (1991). Teachers' instructional actions. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research, Vol. 2 (pp. 861-883). White Plains, NY: Longman.

Roehler, L., Duffy, G., & Warren, S. (1988). Adaptive explanatory actions associated with effective teaching of reading strategies. Dialogues in literacy research: Thirty-seventh yearbook of the National Reading Conference (pp. 339-345). Chicago, IL: National Reading Conference.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: a review of the research. Review of Educational Research, 64(4), 479-530.

Rosenshine, B., & Meister, C. (1997). Cognitive strategy instruction in reading. In S. Stahl & D. Hayes (Eds.), Instructional models in reading. (pp. 85-107). Mahwah, NJ: Lawrence Erlbaum Associates.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. Review of Educational Research, 66(2), 181-221.

Schuder, T. (1993). The genesis of transactional strategies instruction in a reading program for at-risk students. Elementary School Journal, 94(2), 183-200.

Appendix A: Outlines of the Studies

Teacher Participants	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
Number	Total = 22 Treatment group = 11 Control group = 11	Total = 20 Treatment group = 10 Control group = 10	Total = 16 Treatment group = 9 Control group = 7	Total = 10 Treatment group = 5 Control group = 5
Age	Not reported.	Not reported.	Not reported.	Not reported.
Educational background	Not reported.	Not reported.	Not reported.	Not reported.
Years of experience	Not reported.	Not reported.	Not reported.	The SAIL teachers had an average of 10.4 years of general teaching experience, and all of them had taught in the SAIL program for between 3 and 6 years. The comparison group had an average of 23.4 years of teaching experience.
Random assignment to conditions?	Yes. Teachers were observed and given baseline scores on their classroom management skills (high, medium, low). Researchers then randomly assigned teachers within each management level to either the treatment or control group.	Yes.	Yes.	No. SAIL teachers had already been trained before the beginning of the study. The authors state that "preparing teachers to become competent transactional strategies instructors is a long-term process; therefore, we felt we could not randomly assign teachers, provide professional development, and wait for teachers to become experienced in teaching SAIL in a realistic time frame."

Student Participants	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
States represented	Not reported; Midwest USA	One Midwestern state	Not reported.	One mid-Atlantic state
Number of different schools	Not reported.	Treatment group: 9 Control group: 8 All schools were in the same district.	Not reported.	Number not reported; all schools were in the same district.
Number of different classrooms	Total: 22 Treatment group: 11 Control group: 11	Total: 20 Treatment group: 10 Control group: 10	Total: 16 Treatment group: 9 Control group: 7	Total: 10 Treatment group: 5 Control group: 5

Teacher Participants	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
Number	Total = 22 Treatment group = 11 Control group = 11	Total = 20 Treatment group = 10 Control group = 10	Total = 16 Treatment group = 9 Control group = 7	Total = 10 Treatment group = 5 Control group = 5
Age	Not reported.	Not reported.	Not reported.	Not reported.
Educational background	Not reported.	Not reported.	Not reported.	Not reported.
Years of experience	Not reported.	Not reported.	Not reported.	The SAIL teachers had an average of 10.4 years of general teaching experience, and all of them had taught in the SAIL program for between 3 and 6 years. The comparison group had an average of 23.4 years of teaching experience.
Random assignment to conditions?	Yes. Teachers were observed and given baseline scores on their classroom management skills (high, medium, low). Researchers then randomly assigned teachers within each management level to either the treatment or control group.	Yes.	Yes.	No. SAIL teachers had already been trained before the beginning of the study. The authors state that "preparing teachers to become competent transactional strategies instructors is a long-term process; therefore, we felt we could not randomly assign teachers, provide professional development, and wait for teachers to become experienced in teaching SAIL in a realistic time frame."

Student Participants	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
States represented	Not reported, Midwest USA	One Midwestern state	Not reported.	One mid-Atlantic state
Number of different schools	Not reported.	Treatment group: 9 Control group: 8 All schools were in the same district.	Not reported.	Number not reported; all schools were in the same district.
Number of different classrooms	Total: 22 Treatment group: 11 Control group: 11	Total: 20 Treatment group: 10 Control group: 10	Total: 16 Treatment group: 9 Control group: 7	Total: 10 Treatment group: 5 Control group: 5

Number of participants	Total number: not reported. Number per group: ranged from 4 to 22. Average group size = 11.76.	Total: 148 Treatment group: 71 Control group: 77 Number per group: Ranged from 3 to 16 students per class. Overall average: 7.4 per classroom.	Total: 83-Number per group: Ranged from 2 to 10 and was "approximately equal" across groups.	Total: 60 Treatment group: 30 Control group: 30 Number per group: 6
Age	Not reported.	Not reported.	Not reported.	Not reported.
Grade	5th grade.	3rd grade.	Students ranged from 6th through 11th grade.	2nd grade.
Reading level	Low-level reading groups.	Low-level reading groups.	Severely reading disabled.	Reading below 2nd grade level.
Setting	Large urban school district.	Elementary school classrooms in an urban school district in the Midwest.	Not reported.	Unclear.
Exceptional learning characteristics	All students scored more than 1 year below grade level in reading achievement.	"The individuals in the low groups represented the typical range of reading difficulties associated with low-level reading groups in urban centers. Mainstreamed special education students, immigrant children with severe language problems, and students with behavioral disorders were all included."	"All but a very few had been diagnosed as learning disabled," and more than 75% of them had incoming reading levels of grade 3 or below.	None reported.
Selection restrictions	None reported.	None reported.	Not reported.	Only six students in one SAIL class met eligibility requirements, so the researchers decided to use six matched pairs in each classroom as the basis of comparison.
All English speaking?	Yes.	Yes, although the authors note that the sample included "immigrant children with severe language problems."	Yes.	Yes.
Ethnic background	Not reported.	Not reported.	Not reported.	Not reported.

SES	Not reported.	Not reported.	Not reported	Not reported.
Duration of study	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
Total duration of study	One academic year.	One academic year.	Three months.	One academic year.
Number of sessions	Not reported.	Not reported.	Approximately 20.	Not reported.
Minutes per session	Not reported.	Not reported.	30 minutes.	Not reported.
Brief description of instructional approach	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	Direct explanation (DE) with a focus on the use of an explanation model for teaching strategies. The DE approach includes direct explanation of strategy usage, modeling, systematic practice, and scaffolding.	DE with a focus on explaining the reasoning associated with skill and strategy usage. Approach contains all the elements of DE but also requires teachers to analyze the skills prescribed in basal texts and to recast these skills as problemsolving strategies.	TSI with a focus on progressive shifts of teacher attention toward fostering active reading. The TSI approach contains all the elements of DE and also includes extended discussions that emphasize joint construction of text interpretations and student strategy usage.	TSI with a focus on evaluating the effectiveness of an existing TSI program. The TSI approach contains all the elements of DE and also includes extended discussions that emphasize joint construction of text interpretations and student strategy usage.
Specific elements of instructional approach	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
Teacher analysis of skills in basal textbook; - recasting these skills as strategies?-	Yes	Yes	No	No
Direct explanation of strategy usage (What is the strategy? When can it be used? How is it done?)	Yes	Yes	Yes	Yes

Modeling?	Yes	Yes	Yes	Yes
Systematic practice?	Yes	Yes	Yes	Yes
Scaffolding?	Yes	Yes	Yes	Yes
Extended discussions that emphasize joint construction of text interpretations and student strategy usage?	No	No	Yes	Yes
Student choice of reading materials?	No	No	Yes. (Teachers and students collaborated on choice of texts.)	No
Complete description of instructional approach and curricular emphasis	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	The curriculum in this study comprised the skills prescribed in Houghton-Mifflin and Ginn basal textbooks for use with low-level reading groups in the post-primary grades, such as identifying main ideas, drawing conclusions, using glossaries, and decoding. For the purposes of this study, skills are not viewed as rules to be memorized as procedural algorithms. Instead, they are taught as strategies or flexible plans for reasoning about how to remove blockages to meaning. Rather than being applied automatically, skills are applied thoughtfully, consciously, and adaptively.	This research is based on the assertion that "because poor readers lack understanding of the strategic nature of reading, instruction needs to place greater emphasis on the development of poor readers' ability to reason strategically." According to the authors, "it may be necessary when working with poor readers for teachers to explain explicitly, in consistent ways over extended instructional periods, the mental processing associated with [a given] strategy, when it can be used, and how to apply it in a flexible manner." In particular, the authors are interested in the relationship between the explicitness of teacher strategy explanations on the one hand and student strategy awareness and reading ability on the other.	The teacher development model studied in this research is based on the principles of TSI. According to the author, TSI is a method of teaching reading that emphasizes "transactions or negotiations that occur among teacher and students, and students and students while working together to determine text meaning." The view of teacher education presented in this study involves a progressive shift of the teacher's attention. The first stage shifts attention from overt performance of tasks to the underlying comprehension processes. The next stage shifts from teacher questioning, modeling, and explaining to students carrying out these processes. The final stage shifts from students' carrying out active processes under teacher guidance to their assumption of that responsibility themselves.	"The purpose of SAIL is the development of independent, self-regulated meaning-making from text." The SAIL program uses a TSI approach to teaching reading comprehension to low-performing students. According to the authors, "the short-term goal of TSI is the joint construction of reasonable interpretations by group members as they apply strategies to texts. The long-term goal is the internalization and consistently adaptive use of strategic processing whenever students encounter demanding text. Both goals are promoted by teaching reading group members to construct text meaning by emulating expert readers' use of comprehension strategies."

	The particular curricular goal for this study was for readers, when they encounter meaning blockages, to (1) know what skills can be used as strategies for removing the blockage, (2) select a specific strategy, and (3) use that strategy to remove the blockage. Treatment teachers, therefore, were trained to recast basal skills as strategies and to teach students in low-level reading groups to use them when encountering meaning blockages.	Consequently, the instructional approach used in this study focused on teaching students the reasoning that expert readers are presumed to employ when using strategically those skills traditionally taught in association with basal textbooks. Specifically, teachers were taught to recast the skills prescribed in basal textbooks as problem-solving strategies. They were taught to do this by analyzing the cognitive and metacognitive components of the skills and by modeling the cognitive and metacognitive acts involved in performing the skills. The curricular emphasis in the treatment classrooms, therefore, was on the reasoning associated with strategic skill usage, not on the performance of isolated skill tasks.		SAIL teachers are taught to achieve the goals of TSI through direct explanations, Modeling, coaching, and scaffolded practice. In addition, SAIL teachers are taught to facilitate extended discussions of text, which emphasize student application of strategies to text comprehension. In the SAIL reading program, students are taught strategies for adjusting their reading to their specific purpose and to text characteristics. Specifically, students are instructed to predict upcoming events, alter expectations as text unfolds, generate questions and interpretations while reading, visualize represented ideas, summarize periodically, attend selectively to the most important information, and think aloud as they practice applying comprehension strategies during reading instruction. Overreliance on any one strategy is discouraged. In general, students are taught that getting the overall meaning of text is more important than understanding every word.
Materials	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	Basal reading textbooks; difficulty not reported.	Second grade basal reading textbooks.	A total of 135 single-page, expository texts was prepared, and it was left to the teachers and students to decide which of the texts they wished to read.ÑTexts were drawn and edited (primarily shortened) from a variety of "real text" sources (e.g., Cricket Magazine, Open Court Publishing). Readability levels [illegible]nged from grades 2 to 8, with the majority of texts at grades 4 and 5.	It is not entirely clear what texts were used during the course of the school year. The three texts used in the study for assessments were illustrated stories from trade books, with numbers of words and readability levels as follows: 341 words; 2.4 512 words; 2.2 129 words; 3.9

What were treatment teachers taught?	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	Treatment teachers were taught to emphasize the mental processing one does when using the skills prescribed in the basal textbook. The teachers were trained to talk to students about the reasoning one does when encountering a blockage to meaning, how the skill being taught can be applied to remove a particular blockage, and the mental steps one follows when using the skill. Treatment teachers were told to present skills not within the context of workbook exercise but within the context of the use of those skills in actual reading situations. To assist in their planning, teachers were taught to organize their instructional talk into a five-step lesson format: introduction, modeling, guided interaction, practice, and application. To help teachers use the lesson plan, they were taught how to model the mental processing readers do by "talking out loud" about their own use of the skill, direct attention to the salient features of the skill, refocus student attention during interactions, review, provide practice, and help students apply the skill in connected text.	Treatment teachers were taught to modify the curricular and instructional skill prescriptions of the basal text so that the emphasis was on the mental processing involved in using skills as strategies. Specifically, treatment teachers were taught to adapt their basal text instruction in the following ways: Because basal textbooks often present prescribed skills as isolated memory-based tasks, treatment teachers were taught to recast the prescribed skills as problem-solving strategies by analyzing the cognitive and metacognitive components of the skill. Because the teaching suggestions in the basal text teacher's guide emphasize procedural skill exercises and drill, treatment teachers were taught to supplement these suggestions with modeling of the cognitive and metacognitive acts involved in performing the skills. Teachers were taught "to explain explicitly, in consistent ways over extended instructional periods, the mental processing associated with [a given] strategy, when it can be used, and how to apply it in a flexible manner." Teachers were taught "to present their explanations to students as descriptive of what good readers do, rather than as prescriptions to be procedurally applied in all situations." Treatment teachers were not provided with scripts for teaching skills in this way. Instead, they used the information from the research intervention sessions to develop their own explanations for each lesson.	A set of 20 teacher shifts and 12 student shifts was presented to the treatment teachers. The shifts represent changes that need to be made in order to foster more active reading. This list of shifts first describes ways in which teachers and students typically behave in remedial reading sessions, and then provides a contrasting list of behaviors that characterize or promote active reading. The set of student shifts that was presented to teachers included the following as desired goals: Participating in reading to learn new information;- trying to read difficult or unfamiliar material; focusing on collaborating with the group in reading sessions; revealing and investigating errors in reading; directing effort toward explaining how to arrive at correct answers; attempting to take on the role of the teacher; asking questions; reacting to text; providing models for others;-giving elaborated responses; focusing on learning from the reading; and seeking challenges in thinking. Teachers were also given a set of principles for fostering active reading through reading instruction, with specific teacher techniques for each principle. Particular attention was given to: procedures for making thinking explicit by thinking aloud, and for turning over responsibility for this to students, collaborative problem-solving, as well as accessing, applying and evaluating students' existing and alternative problem-solving strategies "upgrading" questioning by both teachers and students to be less content-specific, and more focused on the use of strategies, turning questioning and the entire reading session over to students, and increasing student talk and decreasing teacher talk during reading discussions.	The treatment (SAIL) teachers were not trained specifically for this study; however, they all had extensive experience (i.e., 3 or more years) teaching in the SAIL program.

How did treatment teacher training occur?	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	All teachers attended an initial orientation meeting in November. Subsequent to the initial meeting, the treatment teachers received 10 hours of training on how to incorporate explicit explanations into their ongoing reading skill instruction. This training emphasized: how to recast prescribed basal text skills as strategies useful when removing blockages to meanings, how to make explicit statements about the reading skill being taught, when it would be used, and how to apply it, and how to organize these statements for presentation to students. There were five training sessions, beginning in late November and continuing at about 1-month intervals through March. All the training sessions except one were timed to occur approximately 1 week before each scheduled round of classroom observations. Each training session followed a 4-stage sequence. First, the teachers were provided with information about strategy	Treatment teachers were told that the purpose of the project was to study teacher explanation. They received six 2-hour training sessions in the course of one academic year. These sessions emphasized:-how to make decisions about recasting prescribed basal text skills as strategies; how to decide on explicit statements about the strategy being taught, when it would be used, and how to do the mental processing involved; how to organize these statements into a lesson format that progressed from an introduction, to modeling, to interaction between teacher and students, to closure. The training interventions also included one-on-one coaching, collaborative sharing between the teachers, specific feedback regarding observed lessons, and videotapes of model lessons.	The training of the treatment teachers involved three sessions of 3 hours each, held at one month intervals while the teachers were conducting reading sessions with their students. In these training sessions, treatment teachers were instructed in principles and techniques for fostering active reading. The training module included the following elements and techniques: Research involvement:-The treatment teachers participated in discussions about the study procedures. "Every effort was made to make teachers feel they were a part of the development and evolution of the project." Teaching shifts: As described above, a set of 20 teacher shifts and 12 student shifts, representing changes that need to be made in order for more active reading to be fostered, were presented to the treatment teachers and used throughout their training for self-evaluation. Videotape and self evaluations: At each training session, the teachers were shown videotaped clips of their own teaching and asked to evaluate them in terms of the shifts. During self-evaluation, treatment teachers also discussed and selected the shifts on which they felt they needed the most help and guidance from the experimenter and/or peer teachers. Principles and techniques for fostering active reading: As described above, treatment teachers were given a set of principles for fostering active reading through reading instruction, with specific teacher techniques for each principle. Peer support: Treatment teachers received peer support and	The SAIL teachers were not trained specifically for this study; however, they all had extensive experience (i.e., 3 or more years) teaching in the SAIL program.

	instruction, and links were made to teachers' background experiences in reading instruction, to basal textbook experiences, and to expected student responses. Second, the researchers modeled strategy instruction and assisted teachers as they developed their own instructional plans. Third, teachers read the transcripts of their own previous lessons and student interviews, and the researchers guided them in analyzing and critiquing the transcripts. Finally, the researchers provided teachers with oral feedback following each observation about the appropriateness of their explanations. This feedback was consistent with the information provided to teachers during training interventions.		coaching from previously trained teachers who attended the training sessions and were available as needed for teachers.	
What was the intervention for the control group?	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	The control group received a presentation on effective classroom management. In addition, they were observed teaching classes on four occasions following the baseline observation.	Treated-control teachers were told that the purpose of the study was to validate at the 3rd grade level the results of a previous (unrelated) study involving classroom management for 1st-graders. They received three 2-hour training sessions on using the management principles employed in the 1st grade study. In the classroom, they followed their usual instructional routines regarding basal textbook skill instruction, while adding the management principles of the 1st grade study.	The control teachers were told that they would receive the same training as the treatment teachers after the research data were collected.	The control teachers received no special training; however, they were all "highly regarded for their teaching abilities by district personnel." In addition, the control group had, on average, a greater number of years of teaching experience than the treatment teachers.

What training or information was given to both groups of teachers?	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
	The teachers were unaware that the two groups received different information.	Neither the treatment nor the control group was made aware of the others' existence. Both groups of teachers received identical information about how to implement an Uninterrupted Sustained Silent Reading (USSR) program and how to prepare students to take a standardized reading test.	Not reported.	Not reported.
Outcome measures	**Duffy et al. (1986)**	**Duffy et al. (1987)**	**Anderson (1992)**	**Brown et al. (1996)**
Student reading achievement:	Gates-MacGinitie Reading Test: The comprehension subtest, Level D (designed for use with grades 4-6) was used. (PRE and POST)	Stanford Achievement Test (SAT): The comprehension and word skills subtests were used. (PRE and POST). Michigan Educational Assessment Program (MEAP): The MEAP was administered five months after the treatment ended. (DELAYED POST).	Stanford Diagnostic Reading Test: The phonics, structural analysis, and reading comprehension-subtests were used (PRE and POST).	Stanford Achievement Test (SAT):-The comprehension and word skills subtests were used. (PRE and POST). Story recall: Students were asked cued and picture cued retelling questions about 2 stories. This measure was designed to assess both recall skills and the degree to which students were interpretive in their retelling of the story. (POST).

Student strategy awareness:	Lesson interviews: Immediately following each of the four observed lessons subsequent to the baseline observation, students were interviewed to determine whether they were consciously aware of what strategy the teacher taught during the lesson (declarative knowledge), when to use it (situational knowledge), and how to use it (procedural knowledge). (DURING).	Lesson interviews: Immediately following a reading lesson, students were interviewed to determine whether they were consciously aware of what strategy the teacher taught during the lesson (declarative knowledge), when to use it (situational knowledge), and how to use it (procedural knowledge). (DURING). Concept interviews: At the end of the year, students were interviewed to measure their awareness of the general need to be strategic when reading. (POST).	Not measured.	Strategy awareness interview: In October and November, (i.e., when SAIL components were being introduced to SAIL students) and in March and April, a strategies interview was administered to all students participating in the study. This interview tapped students' reported awareness of strategies, as measured by the number and types of strategies they claimed to use during reading. It was also designed to measure students' awareness of where, when, and why to use strategies. (DURING). Students were asked the following six open-ended questions, adapted from the ones used by Duffy et al. (1987): What do good readers do? What makes someone a good reader? What things do you do before you start to read a story? What do you think about before you start to read a story? What do you do when you come to a word you do not know? What do you do when you read something that does not make sense?
Student strategy usage:	Not measured.	Supplemental Achievement Measure (SAM): This measure was designed by the experimenters to determine whether students could perform the specific skill tasks they had been taught (Part I), and whether their rationale for choosing an answer reflected the reasoning associated with using skills as strategies (Part II). (POST). Modified Graded Oral Reading Paragraph (GORP): This test involved students reading passages orally, and examined self-reports of their self-corrections and their responses to 2 embedded words meeting semantic cueing criteria. (POST).	Not measured.	Think-aloud measure: Students were stopped at four points while reading a difficult story individually with a researcher, and asked to describe their thinking and their strategy usage. (POST).

Teacher effectiveness:	Classroom observation: All teachers in both groups were observed on four separate occasions subsequent to the baseline observation. On the basis of these observations, teachers were rated on the explicitness of their explanations, using a rating scale developed by the researchers. Two aspects of explanation were rated: the information conveyed, and how the teacher conveyed it. The first aspect focused on the content of what the teacher said to students, and was divided into 5 sub-categories: what was said about the skill being taught, when it would be used, the features to attend to, the sequence to follow, and the examples used. The second aspect focused on the pedagogical means by which the information was conveyed, and included 6 sub-categories, focusing on the teachers use of: modeling, highlighting, feedback, review, practice, and application.	Teacher explicitness measure: To measure the explicitness of treatment and treated-control teachers' explanations, the researchers developed an instrument to rate transcripts of audiotaped lessons. (DURING). The rating instrument was organized into three parts: Part I of the instrument focused on the information presented. Teachers were rated on what they students said to students about (a) the task to be learned, (b) its usefulness, (c) the selection of the strategy to be used, and (d) how to do the mental processing associated with the strategy). Part II focused on the means used to present information. Teachers were rated on their (a) introduction to the lesson, (b) modeling, (c) diminishing assistance during interaction, (d) eliciting of student responses, and (e) closure.ÑPart III focused on the cohesion both within the lesson and across lessons.Ñ	Videotaped classroom observation: Teachers were videotaped giving a reading lesson for approximately 30 minutes. (PRE and POST). A rating scale was developed using the teacher and student shifts as a base. Teachers were rated on the following 14 dimensions: Treating reading problems openly, focusing on how to solve problems, providing models of thinking, teaching question-asking, asking thought-provoking questions, allowing student control, focusing on group collaboration, informing students of learning, focusing on text and learning about reading, setting reading goals before reading, problem solving during reading, summarizing to check comprehension, reflecting on reading goals after text, and stressing new learning from text. Teacher effectiveness was also assessed by rating students on the following 8 dimensions: Treating reading problems openly, focusing on how to solve problems, expressing thinking, asking questions, giving elaborated answers, taking teacher role, focusing on group collaboration, involvement in sessions.	Classroom observation: SAIL and non-SAIL teachers were observed teaching two story lessons and were compared in terms of the number of strategies they taught in each lesson. (DURING).

Results	Duffy et al. (1986)	Duffy et al. (1987)	Anderson (1992)	Brown et al. (1996)
Student reading achievement:	Gates-MacGinitie Reading Test: There was no significant difference between students in the treatment and control classrooms on the comprehension subtest at posttest (ES = 0.24). Students in treatment and control classrooms spent equal amounts answering comprehension test items on the pretest, but on the posttest, treatment students spent significantly more time answering questions (ES = +0.42).	Stanford Achievement Test (SAT):-Students of treatment teachers scored significantly higher than students of control teachers on word skills (ES = +1.63), but not on comprehension (ES = +0.25). Michigan Educational Assessment Program (MEAP): Students of treatment teachers scored significantly higher than students of control teachers (ES = +1.33).	Stanford Diagnostic Reading Test: A significantly higher number of students of treatment teachers (about 80%) made gains on the reading comprehension subtest than students of control teachers (about 50%). There was no significant difference in the number of students of treatment teachers and the number of students of control teachers who made gains on the phonics and the structural analysis subtests.	Stanford Achievement Test (SAT): Students of treatment teachers scored significantly higher than students of control teachers on the comprehension subtest (ES = +1.70) and the word skills subtest (ES = +1.67); they also showed significantly greater improvement on these measures over the course of the study.ÑStory retelling questions: Students of the treatment (SAIL) teachers recalled more literal information (Story 1: ES = +0.69; Story 2: ES = +1.37) and were significantly more interpretive in their retelling of the stories (Story 1: ES = +1.01; Story 2: ES = +1.07) than were students of control teachers.
Student strategy awareness:	Strategy awareness interview: Students of treatment teachers scored significantly higher than students of control teachers on strategy awareness ratings (ES = +1.39).	Lesson interviews: Lesson interview responses of students of treatment teachers were rated significantly higher than the responses of students of control teachers. These findings were due to significantly higher ratings given to students of treatment teachers on situational knowledge (ES = +2.22) and procedural knowledge (ES = +1.50). No difference in response ratings was found between groups for declarative knowledge (ES = +0.84). Concept interviews:-Concept interview responses of students of the treatment teachers were rated significantly higher than the responses of students of the control teachers (ES = +1.15), thus suggesting that the treatment students were more aware of the strategic nature of reading.	Not measured.	Strategies interview: Toward the end of the treatment, the students of the treatment (SAIL) teachers reported more awareness of comprehension (ES = +4.03) and word-level strategies (ES = +1.38) during the interview than did the students of control group teachers.

Student strategy usage	Not measured.	*SAM Test:* Students of treatment teachers did not differ significantly from students of control teachers in their performance on Part I (ES = -0.21). However, students of treatment teachers were significantly superior to students of control teachers in their performance on Part II (ES = +1.67). Modified GORP Test: -Students of treatment teachers scored significantly higher on both the word meaning subtest (ES = +1.51) and the word recognition subtest (ES = +5.00). "According to these GORP results, low-group students who received explicit explanations about the reasoning associated with using skills as strategies (a) reported that they used such reasoning when actually reading connected text, and (b) described the reasoning employed when using the strategies."	*Not measured*	Think-aloud measure: Students of treatment (SAIL) teachers applied significantly more strategies during the think-aloud task than did the students of control teachers (ES = +2.98).
Teacher effectiveness	Teacher explicitness measure: Across all observations after the baseline observation, treatment teachers were rated as significantly more explicit in their explanations than control teachers (ES = +2.11).	Teacher explicitness measure: The treatment teachers were found to be more explicit in explaining the reasoning associated with using reading skills as strategies than the treated-control teachers (ES = +1.67).	Videotaped teaching sessions: The treatment teachers showed significant improvements across all 14 dimensions: Treating reading problems openly (ES = +3.80), focusing on how to solve problems (ES = +2.80),-providing models of thinking (ES = +3.25), teaching question-asking (ES = +2.00), asking thought-provoking questions (ES = +3.14),-allowing student control (ES = +2.08), focusing on group collaboration (ES = +2.56),-informing students of learning (ES = +2.35), focusing on text and learning about	Classroom observations: The treatment (SAIL) teachers were found to have taught significantly more comprehension strategies (ES = +5.48) and more word-level strategies (ES = +1.38) than control teachers.

			reading (ES = +2.52), setting reading goals before reading (ES = +3.99), problem solving during reading (ES = +5.73), summarizing to check comprehension (ES = +1.90), reflecting on reading goals after text (ES = +2.21), and stressing new learning from text (ES = +2.45). Students of treatment teachers showed significant improvement on all 8 dimensions: Treating reading problems openly (ES = +3.24), focusing on how to solve problems (ES = +3.20), expressing thinking (ES = +2.85), asking questions (ES = +2.81), giving elaborated answers (ES = +1.48), taking teacher role (ES = +2.74), focusing on group collaboration (ES = +2.46), and involvement in sessions (ES = +2.14). Treatment teachers showed a far greater percentage of problem solving incidents at posttest than at pretest. No statistical test is presented. There was a significant increase in student talk and a decrease in teacher talk in the treatment condition. The relevant data are not presented. "It is clear... that the experimental teachers and their students changed substantially from pre- to posttest, while control teachers and students remained about the same."	

Appendix B:

Comprehensive Summaries Based on NRP Guidelines

Duffy et al. (1986)

1. Reference

Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J. & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. Reading Research Quarterly, 21(3), 237-252.

2. Research Question

The goal of this study was to determine whether, given skills prescribed in a mandated basal reading series, classroom teachers of low-group students who provide more explicit explanations of how to use these reading skills strategically would be more effective than teachers who were less explicit in explaining how to use skills.

The authors hypothesized that explicit teacher explanation would result in improved student awareness about what was taught, which in turn would result in increased reading achievement on a standardized measure.

The study sought to answer the following questions:

- Are teachers trained to be more explicit during low-group reading skill instruction more explicit than teachers who receive no training?
- Are low-group students of teachers who receive training in how to provide explicit explanation more aware of what skill was taught and of how to use it strategically than low-group students of teachers who receive no training?
- Do the low-group students of trained teachers score significantly higher on the comprehension subtest of a standardized reading achievement test than low-group students of untrained teachers?

3. Sample of student participants

States or countries represented: Not reported. Midwest, USA

Number of different schools: Not reported.

Number of different classrooms:
- *Total:* 22
- *Treatment group:* 11
- *Control group:* 11

Number of participants (total, per group):
- *Total number:* Not reported.
- *Number per group:* Ranged from 4 to 22.
- *Average group size* = 11.76.

Age: Not reported

Grade: 5th.

Reading levels of participants: Low reading groups.

Setting: Large urban school district.

Pretests administered prior to treatment:
- Form 2 of the Gates-MacGinitie was administered in early October to low-group students in all 22 classrooms.

Special characteristics:
- *SES:* Not reported.
- *Ethnicity:* Not reported.
- *Exceptional learning characteristics:*
  - *Learning disabled:* Not reported.
  - *Reading disabled:* Not reported.
  - *Hearing impaired:* Not reported.
  - *English language learners (LEP):* Not reported.

Selection restrictions used to limit the sample of participants: None reported.

Contextual information (concurrent reading instruction that participants received in their classrooms during the study): Not reported.

Description of curriculum/instructional approach:

Direct explanation (DE) with a focus on the use of an explanation model for teaching strategies. The DE approach includes direct explanation of strategy usage, modeling, systematic practice, and scaffolding.

The curriculum in this study comprised the skills prescribed in the Houghton-Mifflin and Ginn basal textbooks for use with low reading groups in the postprimary grades, such as identifying main ideas, drawing conclusions, using glossaries, and decoding. For the purposes of this study, skills are not viewed as rules to be memorized as procedural algorithms. Instead, they are taught as strategies, or flexible plans for reasoning about how to remove blockages to meaning. Rather than being applied automatically, skills are applied thoughtfully, consciously, and adaptively.

The recasting of traditional reading skills as strategies is based on cognitive science research and on the application of such research to reading comprehension.

The particular curricular goal for this study was for readers, when they encounter meaning blockages, to (a) know what skills can be used as strategies for removing the blockage, (b) select a specific strategy, and (c) use that strategy to remove the blockage.

Treatment teachers, therefore, were trained to recast basal skills as strategies and to teach students in low reading groups to use them when encountering meaning blockages.

How was the sample obtained?

The teachers volunteered in response to a survey of all 5th grade teachers of low reading groups in the district. The students were assigned to reading groups by teachers as part of the participating school district's policy of using the Joplin Plan to group 5th grade students homogeneously for reading. Student assignments to reading groups were made on the basis of Stanford Achievement Test scores from the previous year and the recommendations of previous teachers. All the low-group students in this study scored more than 1 year below grade level in reading achievement.

Attrition: Not reported.

4. Setting of the Study

Elementary school classroom with low-group reading students.

5. Design of the Study

Random assignment of participants (teachers) to treatments (randomized experiment), after a pretest of classroom management skills and stratification on this dimension.

6. Independent Variables

a. Treatment variables

Describe all treatments and control conditions.

All teachers attended an initial orientation meeting in November. Subsequent to the initial meeting, the treatment teachers received 10 hours of training on how to incorporate explicit explanations into their ongoing reading skill instruction. This training emphasized:

- How to recast prescribed basal text skills as strategies useful when removing blockages to meanings
- How to make explicit statements about the reading skill being taught, when it would be used and how to apply it
- How to organize these statements for presentation to students.

Specifically, treatment teachers were taught to emphasize the mental processing one does when using the skills prescribed in the basal textbook. The teachers were trained to talk to students about

- The reasoning one does when encountering a blockage to meaning
- How the skill being taught can be applied to remove a particular blockage
- The mental steps one follows when using the skill.

That is, teachers were told to present skills not within the context of workbook exercises but within the context of the use of those skills in actual reading situations.

To assist in their planning, teachers were taught to organize their instructional talk into a five-step lesson format: introduction, modeling, guided interaction, practice, and application. To help teachers use the lesson plan, they were taught how to

- Model the mental processing readers do by "talking out loud" about their own use of the skill

- Direct attention to the salient features of the skill and how to refocus student attention during interactions
- Review
- Provide practice
- help students apply the skill in connected text.

The five training sessions were conducted immediately after school, beginning in late November and continuing at about 1-month intervals through March. All the training sessions except one were timed to occur approximately 1 week before each scheduled round of classroom observations.

Each training session followed a four-stage sequence. First, the teachers were provided with information about strategy instruction, and links were made to teachers' background experiences in reading instruction, to basal textbook experiences, and to expected student responses. Second, the researchers modeled strategy instruction and assisted teachers as they developed their own instructional plans. Third, teachers read the transcripts of their own previous lessons and student interviews, and the researchers guided them in analyzing and critiquing the transcripts. Finally, the researchers provided teachers with oral feedback following each observation about the appropriateness of their explanations. This feedback was consistent with the information provided to teachers during training interventions.

The control group received a presentation on effective classroom management. In addition, these teachers were observed teaching classes on four occasions following the baseline observation.

Was instruction explicit or implicit? Explicit.

Difficulty and nature of texts used: Basal texts, difficulty not reported.

Was trainers' fidelity in delivering treatment checked? Yes, via classroom observation.

Properties of trainers (teachers)

Number of teachers who administered treatments:

Experimental = 11
Control = 11
Total = 22

Teacher/student ratio: Not reported

Type of trainer (teacher): Classroom teacher

Length of training given to trainers (teachers): See above.

Source of training: The researchers.

Assignment of trainers (teachers) to group:

Teachers were observed and given baseline scores on their classroom management skills (high, medium, low). This resulted in teachers being assigned to the following management levels:

"High" = 8
"Average" = 4
"Low" = 2

Researchers then randomly assigned teachers within each management level to either the treatment or control group.

Management ratings were made again at four observation points during the year to validate the initial management ratings.

Teachers were also observed at the beginning of the study to obtain a baseline measure of their skill instruction to establish that all 22 teachers were relatively equal in the explicitness of their explanations.

Baseline data were unavailable for two teachers (1 treatment and 1 control).

Cost factors: Not reported.

b. Moderator variables

List and describe other nontreatment independent variables included in the analyses of effects: None reported.

7. Dependent (Outcome) Variables

List processes that were taught during training and measured during and at the end of training: See #6 above.

Student strategy awareness:

Student awareness data for both treatment and control classrooms were obtained in interviews with five randomly selected low-group students from each classroom immediately following each of the four observed lessons subsequent to the baseline observation. The same five students were interviewed each time, except in the case of one classroom that had only four low-group students, where all four were interviewed each time. If a designated student was absent or moved away during the study, another student from the low-reading group was randomly selected to complete the complement of five interviewees.

Three questions were asked of each student, followed by prepared probes if responses to the initial questions were incomplete or vague.

- What were you learning in the lesson I just saw?
- When would you use what was taught in the lesson?
- How do you do what you were taught to do?

The criteria for determining student awareness were contained in a rating scale developed by the research team. Ratings ranged from 0 to 4 on each of the following three criteria:

1. Awareness of what had been taught
2. Awareness of the context or situation in which the strategy should be used or applied
3. Awareness of how the strategy is employed.

Teacher explicitness:

All teachers in both groups were observed on four separate occasions subsequent to the baseline observation. On the basis of these observations, teachers were rated on the explicitness of their explanations, using a rating scale developed by the researchers. Two aspects of explanation were rated: the information conveyed and how the teacher conveyed it.

The first aspect focused on the content of what the teacher said to students and was divided into five subcategories:

1. What was said about the skill being taught
2. When it would be used
3. The features to attend to
4. The sequence to follow
5. The examples used.

The second aspect focused on the pedagogical means by which the information was conveyed and included six subcategories, focusing on the teachers' use of:

- Modeling
- Highlighting
- Feedback
- Review
- Practice
- Application.

Teachers received ratings for degrees of explicitness on each of the 11 subcategories on a scale of 0 to 2 (with 0 indicating absence, and 2, exemplary presence of the criterion).

Student Achievement

The achievement measure was the comprehension subtest of the Gates-MacGinitie Reading Test (2nd ed., MacGinitie, 1978), Level D (designed for use with grades 4 to 6). This test consists of short paragraphs followed by a series of two to four multiple-choice questions about the content of each paragraph (43 total items). Form 2 was given as the pretest and Form 1 as the posttest.

8. Nonequivalence of groups

Any reason to believe that treatment and control groups might not have been equivalent prior to treatments?

No. "Although baseline data were not available for student awareness ratings, the stratified random assignment of teachers to treatment and control groups, coupled with the similarity of baseline explanation ratings (4.1 for each group) and the similarity of pretest comprehension scores, suggests that there was no initial awareness [or achievement] difference between groups."

Were steps taken in statistical analyses to adjust for any lack of equivalence?

Not reported.

9. Result (for each measure)

a. Name of Measure: Student strategy awareness interview

Students of treatment teachers scored significantly higher than students of control teachers on strategy awareness ratings.

Value of effect size: +1.39

Type of summary statistics from which effect size was derived: ANOVA

Number of classrooms providing the effect size information: Ns = 11 and 11

b. Name of Measure: Teacher explicitness

Across all observations after the baseline observation, treatment teachers were rated as significantly more explicit in their explanations than control teachers.

Value of effect size: +2.11.

Type of summary statistics from which effect size was derived: ANOVA

Number of classrooms providing the effect size information: *Ns = 11 and 11*

c. Name of Measure: Student Achievement

There was no significant difference between students in the treatment and control classrooms on the comprehension subtest at either pretest or posttest.

Value of effect size: 0.24.

Type of summary statistics from which effect size was derived: ANOVA

Number of classrooms providing the effect size information: Ns = 11 and 11

Students in treatment and control classrooms spent equal amounts of time answering comprehension test items on the pretest, but on the posttest, treatment students spent significantly more time answering questions.

Value of effect size: +0.42

Type of summary statistics from which effect size was derived: t-test.

Number of classrooms providing the effect size information: Ns = 11 and 11

Duffy et al. (1987)

1. Reference

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M.S., Vavrus, L. G., Wesselman, R., Putname, J. & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. Reading Research Quarterly, 23(3), 347-368.

2. Research Question

The purpose of this study was to investigate the effects of explaining the reasoning associated with using reading strategies. Three specific research questions were posed.

- Can teachers learn to be more explicit in explaining the reasoning associated with using basal text skills as strategies?
- Can explicit teacher explanations increase low-group students' awareness of both lesson content and the need to be strategic while reading?
- Can explicit teacher explanations increase low-group students' conscious use of skills as strategies and lead, ultimately, to greater reading achievement?

3. Sample of student participants

States or countries represented: The Midwest (no state given), USA

Number of different schools: Treatment Group = 8; Control Group = 9

Number of different classrooms = 20

Number of student participants:

Total: 148

Treatment group: 71

Control group: 77

Number per group: Ranged from 3 to 16 students per class.

Overall average: 7.4 per classroom.

Age: Not listed

Grade: 3rd

Reading levels of participants: "Low"

Setting: Urban, suburban

Pretests administered prior to treatment:

Stanford Achievement Test (SAT), reading section, administered at end of 2nd grade.

Special characteristics, if relevant:

SES: Not reported.

Ethnicity: Not reported.

Exceptional Learning Characteristics:

These students "represented the typical range of reading difficulties associated with low reading groups in urban centers." Groups included mainstreamed special education students, immigrant children with severe language problems, and students with behavioral disorders.

Selection restrictions used to limit the sample of participants: Not reported.

Contextual information (concurrent reading instruction that participants received in their classrooms during the study): Not reported.

Description of curriculum/instructional approach:

Direct explanation (DE) with a focus on explaining the reasoning associated with skill and strategy usage.

Duffy et al.'s approach contains all the elements of DE but also requires teachers to analyze the skills prescribed in basal texts, and to recast these skills as problem-solving strategies.

This research is based upon the assertion that "because poor readers lack understanding of the strategic nature of reading, instruction needs to place greater emphasis on the development of poor readers' ability to reason strategically."

According to the authors, "it may be necessary when working with poor readers for teachers to explain explicitly, in consistent ways over extended instructional periods, the mental processing associated with [a given] strategy, when it can be used, and how to apply it in a flexible manner."

In particular, the authors are interested in the relationship between the explicitness of teacher strategy explanations on the one hand and student strategy awareness and reading ability on the other.

Consequently, the instructional approach used in this study focused on teaching students the reasoning that expert readers are presumed to employ when using strategically those skills traditionally taught in association with basal textbooks.

Specifically, teachers were taught to recast the skills prescribed in basal textbooks as problem-solving strategies. They were taught to do this by analyzing the cognitive and metacognitive components of the skills, and by modeling the cognitive and metacognitive acts involved in performing the skills.

The curricular emphasis in the treatment classrooms, therefore, was on the reasoning associated with strategic skill usage, not on the performance of isolated skill tasks.

How sample was obtained: Selected from the population of those available.

Attrition: One urban teacher was replaced by a suburban teacher in mid-September.

4. Setting of the Study

Classrooms for low-level reading groups.

5. Design of the Study

Random assignment of participants (teachers) to treatments (randomized experiment). Each teacher's pre-existing reading groups remained intact. Pretest measures revealed no significant differences between the participating groups of students.

6. Independent Variables

a. Treatment variables

Describe all treatments and control conditions.

Treatment teachers were taught to modify the curricular and instructional skill prescriptions of the basal text so that the emphasis was on the mental processing involved in using skills as strategies. Specifically, treatment teachers were taught to adapt their basal text instruction in the following ways:

- Because basal textbooks often present prescribed skills as isolated memory-based tasks, treatment teachers were taught to recast the prescribed skills as problem-solving strategies by analyzing the cognitive and metacognitive components of the skill.
- Because the teaching suggestions in the basal text teacher's guide emphasize procedural skill exercises and drill, treatment teachers were taught to supplement these suggestions with modeling of the cognitive and metacognitive acts involved in performing the skills.
- Teachers were taught "to explain explicitly, in consistent ways over extended instructional periods, the mental processing associated with [a given] strategy, when it can be used, and how to apply it in a flexible manner."
- Teachers were taught "to present their explanations to students as descriptive of what good readers do, rather than as prescriptions to be procedurally applied in all situations."

Treatment teachers were not provided with scripts for teaching skills in this way. Instead, they used the information from the research intervention sessions to develop their own explanations for each lesson.

Treatment teachers were told that the purpose of the project was to study teacher explanation. They received six 2-hour training sessions in the course of one academic year. These sessions emphasized how to

- Make decisions about recasting prescribed basal text skills as strategies.
- Decide on explicit statements about the strategy being taught, when it would be used, and how to do the mental processing involved.
- Organize these statements into a lesson format that progressed from an introduction, to modeling, to interaction between teacher and students, to closure.

The training interventions also included one-on-one coaching, collaborative sharing between the teachers, specific feedback regarding observed lessons, and videotapes of model lessons.

Treated-control teachers were told that the purpose of the study was to validate at the 3rd grade level the results of a previous (unrelated) study involving classroom management for 1st graders. They received three 2-hour training sessions on using the management principles employed in the 1st grade study. In the classroom, they followed their usual instructional routines regarding basal textbook skill instruction, while adding the management principles of the 1st grade study.

Neither the treatment nor the control group was made aware of the other's existence.

Both groups of teachers received identical information about how to implement an uninterrupted sustained silent reading (USSR) program and how to prepare students to take a standardized reading test.

Was instruction explicit or implicit? Explicit.

Difficulty and nature of texts used: Basal reading textbooks for the 2nd grade.

Was trainers' (teachers') fidelity in delivering treatment checked? Yes, by observations and checklists.

Properties of teachers/trainers:

Number of teachers who administered treatments:

Treatment group = 10

Control group = 10

Total = 20

Teacher/student ratio: Depended on class; ranged from 1:3 to 1:16.

Type of trainer: Classroom teacher.

Any special qualification of trainers (teachers)? No.

Length of training given to trainers (teachers): 12 hours (six 2-hour sessions over the course of the school year).

Source of training: The researchers.

Assignment of trainers to groups: Teachers were already assigned to students at beginning of study.

Cost factors: Not reported.

b. Moderator variables:

List and describe other nontreatment independent variables included in the analyses of effects: None reported.

7. Dependent (Outcome) Variables

Student reading achievement:

Stanford Achievement Test (SAT):
The comprehension and word skills subtests were used. (PRE and POST).

Michigan Educational Assessment Program (MEAP):
The MEAP was administered 5 months after the treatment ended.
(DELAYED POST).

Student strategy awareness:

Lesson interviews:
Immediately following a reading lesson, students were interviewed to determine whether they were consciously aware of what strategy the teacher taught during the lesson (declarative knowledge), when to use it (situational knowledge), and how to use it (procedural knowledge).

(DURING).

Concept interviews:

At the end of the year, students were interviewed to measure their awareness of the general need to be strategic when reading.

(POST).

Student strategy usage:

Supplemental Achievement Measure (SAM):

This measure was designed by the experimenters to determine whether students could perform the specific skill tasks they had been taught (Part I) and whether their rationale for choosing an answer reflected the reasoning associated with using skills as strategies (Part II).

(POST).

Modified Graded Oral Reading Paragraph (GORP):

This test involved students reading passages orally and examined self-reports of their self-corrections and their responses to two embedded words meeting semantic cueing criteria.

(POST).

Teacher effectiveness:

Teacher explicitness measure:

To measure the explicitness of treatment and treated-control teachers' explanations, the researchers developed an instrument to rate transcripts of audiotaped lessons.

(DURING).

The rating instrument was organized into three parts:

- Part I of the instrument focused on the information presented. Teachers were rated on what they said to students about (1) the task to be learned, (2) its usefulness, (3) the selection of the strategy to be used, and (4) how to do the mental processing associated with the strategy.
- Part II focused on the means used to present information. Teachers were rated on their (1) introduction to the lesson, (2) modeling, (3) diminishing assistance during interaction, (4) eliciting of student responses, and (5) closure.
- Part III focused on the cohesion both within the lesson and across lessons.

8. Nonequivalence of groups

Any reason to believe that treatment and control groups might not have been equivalent before treatments? No.

Were steps taken in statistical analyses to adjust for any lack of equivalence? Yes.

9. Result (for each measure):

Student reading achievement:

a. Name of measure: SAT: Word Skills

Students of treatment teachers scored significantly higher than students of control teachers on word skills.

Value of effect size: +1.63

Type of summary statistics from which effect size was derived:

MANCOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

b. Name of measure: SAT: Comprehension

Students of treatment teachers did not score significantly higher than students of control teachers on comprehension.

Value of effect size: +0.25

Type of summary statistics from which effect size was derived:

MANCOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

c. Name of measure: MEAP

Students of treatment teachers scored significantly higher than students of control teachers.

Value of effect size: +1.33

Type of summary statistics from which effect size was derived:

ANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

Student strategy awareness:

d. Name of measure: Lesson interviews

Lesson interview responses of students of treatment teachers were rated significantly higher than the responses of students of control teachers. These findings were due to significantly higher ratings given to students of treatment teachers on situational knowledge and procedural knowledge.

Value of effect size:

Declarative knowledge: +0.84

Situational knowledge: +2.22

Procedural knowledge: +1.50

Type of summary statistics from which effect size was derived:

ANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

e. Name of measure: Concept interviews

Concept interview responses of students of the treatment teachers were rated significantly higher than the responses of students of the control teachers, thus suggesting that the treatment students were more aware of the strategic nature of reading.

Value of effect size: +1.15

Type of summary statistics from which effect size was derived:

MANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

f. Student strategy usage:

Name of measure: SAM: Part II

(performance of skill tasks)

Students of treatment teachers did not differ significantly from students of control teachers in their performance on Part I.

Value of effect size: -0.21

Type of summary statistics from which effect size was derived:

MANOVA

g. Number of classrooms providing the effect size information:

Ns = 10 and 10

Name of measure: SAM: Part II

(reasoning associated with use of skills as strategies)

Students of treatment teachers were significantly superior to students of control teachers in their performance on Part II.

Value of effect size: +1.67

Type of summary statistics from which effect size was derived:

MANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

h. Name of measure:

GORP: Word meaning ratings

Students of treatment teachers scored significantly higher on the word meaning subtest.

Value of effect size: +1.51

Type of summary statistics from which effect size was derived:

MANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

i. Name of measure:

GORP: Word recognition ratings

Students of treatment teachers scored significantly higher on the word recognition subtest.

Value of effect size: +5.00

Type of summary statistics from which effect size was derived:

MANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

"According to these GORP results, low-group students who received explicit explanations about the reasoning associated with using skills as strategies (1) reported that they used such reasoning when actually reading connected text and (2) described the reasoning employed when using the strategies."

Teacher effectiveness:

j. Name of measure: Teacher explicitness measure

The treatment teachers were found to be more explicit in explaining the reasoning associated with using reading skills as strategies than the treated-control teachers.

Value of effect size: +1.67

Type of summary statistics from which effect size was derived:

ANOVA

Number of classrooms providing the effect size information:

Ns = 10 and 10

Anderson (1992)

1. Reference

Anderson, V. (1992). A teacher development project in transactional strategy instruction for teachers of severely reading-disabled adolescents. Teaching and Teacher Education, 8(4) 391-403.

2. Research Question

The purpose of this study was to test the effectiveness of a teacher development model designed to provide teachers with collaborative transactional strategies for helping severely reading-delayed adolescents take a more active approach to understanding informational texts.

The research question addressed by this study is: Does the use of the TSI approach to reading instruction result in positive changes in severely reading-delayed adolescent students' reading performance?

3. Sample of Student Participants

States or countries represented: Not reported.

Number of different schools: Not reported.

Number of participants (total, per group):

Total: 83

Per group: Ranged from 2 to 10 and was "approximately equal" across groups.

Age: Not reported.

Grade: Ranged from 6 to 11.

Reading levels of participants:

Severely reading disabled: "All but a very few had been diagnosed as learning disabled." More than 75% of the adolescent students in the study had incoming reading levels of grade 3 or below.

Setting: Not reported.

Pretests administered before to treatment:

At the beginning of the study, teachers in both an experimental and a control group were videotaped giving a reading lesson for approximately 30 minutes, using one of two expository passages developed for the study that were matched for difficulty but had different content.

In addition, students were given three subtests of the Stanford Diagnostic Reading Test (phonics, structural analysis, and reading comprehension).

The purpose of these steps was to establish pretest baseline measures of teaching style and student ability.

Special characteristics, if relevant:

SES: Not reported.

Ethnicity: Not reported.

Exceptional learning characteristics?

Learning disabled: yes

Reading disabled: yes

Selection restrictions used to limit the sample of participants: Not reported.

Contextual information (concurrent reading instruction that participants received in their classrooms during the study): Not reported.

Description of curriculum/instructional approach:

TSI with a focus on progressive shifts of teacher attention toward fostering active reading. The TSI approach contains all the elements of DE and also includes extended discussions that emphasize joint construction of text interpretations and student strategy usage.

According to the author, TSI is a method of teaching reading that emphasizes "transactions or negotiations that occur among teacher and students, and students and students while working together to determine text meaning."

The view of teacher education presented in this study involves a progressive shift of the teacher's attention.

- The first stage is to shift the attention from overt performance of tasks to the underlying comprehension processes.
- The next stage shifts from teacher questioning, modeling, and explaining to students carrying out these processes.
- The final stage shifts from students' carrying out active processes under teacher guidance to their assuming that responsibility themselves.

How was sample obtained?

Teachers were invited to volunteer via a letter from the participating board of education.

Attrition:

Experimental: 1 teacher

Control: 3 teachers

(originally, there were 10 teachers in each group)

4. Setting of the Study

Small-group reading session, in which teachers work directly with students on the reading and understanding of informational text.

5. Design of the Study

Random assignment of participants (teachers) to treatments (randomized experiment).

6. Independent Variables

a. Treatment variables:

Describe all treatments and control conditions:

A set of 20 teacher shifts and 12 student shifts was presented to the treatment teachers. The shifts represent changes that need to be made for more active reading to be fostered. This list of shifts first describes ways in which teachers and students typically behave in remedial reading sessions; it then provides a contrasting list of behaviors that characterize or promote active reading. The set of student shifts that was presented to teachers included the following as desired goals:

- Participating in reading to learn new information
- Trying to read difficult or unfamiliar material
- Focusing on collaborating with the group in reading sessions
- Revealing and investigating errors in reading
- Directing effort toward explaining how to arrive at right answers
- Attempting to take on the role of the teacher
- Asking questions
- Reacting to text
- Providing models for others
- Giving elaborated responses
- Focusing on learning from the reading
- Seeking challenges in thinking.

Teachers were also given a set of principles for fostering active reading through reading instruction, with specific teacher techniques for each principle. Particular attention was given to:

- Procedures for making thinking explicit by thinking aloud and for turning over responsibility for this to students
- Collaborative problem-solving, as well as accessing, applying, and evaluating students' existing and alternative problem-solving strategies
- "Upgrading" questioning by both teachers and students to be less content-specific and more focused on the use of strategies
- Turning questioning and the entire reading session over to students and increasing student talk and decreasing teacher talk during reading discussions.

The training of the treatment teachers involved three sessions of 3 hours each, held at 1-month intervals while the teachers were conducting reading sessions with their students. In these training sessions, treatment teachers were instructed in principles and techniques for fostering active reading. The training module included the following elements and techniques:

Research involvement:

The treatment teachers participated in discussions about the study procedures. "Every effort was made to make teachers feel they were a part of the development and evolution of the project."

Teaching shifts:

As described above, a set of 20 teacher shifts and 12 student shifts, representing changes that need to be made in order for more active reading to be fostered, was presented to the treatment teachers and used throughout their training for self-evaluation.

Videotape and self-evaluations:

At each training session, the teachers were shown videotaped clips of their own teaching and asked to evaluate them in terms of the shifts. During self-evaluation, treatment teachers also discussed and selected the shifts on which they felt they needed the most help and guidance from the experimenter and/or peer teachers.

Principles and techniques for fostering active reading:

As described above, treatment teachers were given a set of principles for fostering active reading through reading instruction, with specific teacher techniques for each principle.

Peer support:

Treatment teachers received peer support and coaching from previously trained teachers who attended the training sessions and were available as needed for teachers.

The control teachers were told that they would receive the same training as the treatment teachers after the research data were collected.

Was instruction explicit or implicit? Explicit.

Difficulty and nature of texts used:

A total of 135 single-page, expository texts was prepared, and it was left to the teachers and students to decide which of the texts they wished to read during the approximately 20 reading sessions in which they would engage.

Texts were drawn and edited (primarily shortened) from a variety of "real text" sources, e.g., Cricket Magazine.

Readability levels ranged from grades 2 to 8, with the majority of texts at grades 4 and 5.

(Because the intervention included a particular emphasis on identifying reading problems and sharing problem-solving strategies, all texts were somewhat challenging so that problems would arise during reading.)

Was trainers' (teachers') fidelity in delivering treatment checked?

Yes: experimental teachers were videotaped 3 times during the study (pretest, intervention, and posttest).

Properties of trainers (teachers):

Number of trainers (teachers) who administered treatments:

Experimental: 9

Control: 7

Total: 16

Teacher/student ratio: Not reported.

Type of trainer (teacher): Classroom teacher.

Any special qualification of trainers?

All of the teachers were experienced special education teachers.

Length of training given to teachers:

Experimental teachers participated in three afternoon sessions of 3 hours each, held at 1-month intervals while the teachers were conducting reading sessions with their students.

Source of training: The researchers

Assignment of trainers to groups: Random

Cost factors: Not reported.

b. Moderator variables:

List and describe other nontreatment independent variables included in the analyses of effects: None reported.

7. Dependent (Outcome) Variables

Student reading achievement:

Stanford Diagnostic Reading Test:

The phonics, structural analysis, and reading comprehension

subtests were used.

(PRE and POST).

Teacher effectiveness:

Videotaped classroom observation:

Teachers were videotaped giving a reading lesson for approximately 30 minutes. (PRE and POST). A rating scale was developed using the teacher and student shifts as a base. Teachers were rated on the following 14 dimensions:

1. Treating reading problems openly
2. Focusing on how to solve problems
3. Providing models of thinking
4. Teaching question-asking
5. Asking thought-provoking questions
6. Allowing student control
7. Focusing on group collaboration
8. Informing students of learning
9. Focusing on text and learning about reading

10. Setting reading goals before reading
11. Problem-solving during reading
12. Summarizing to check comprehension
13. Reflecting on reading goals after text
14. Stressing new learning from text.

Teacher effectiveness was also assessed by rating students on the following eight dimensions:

1. Treating reading problems openly
2. Focusing on how to solve problems
3. Expressing thinking
4. Asking questions
5. Giving elaborated answers
6. Taking teacher role
7. Focusing on group collaboration
8. Involvement in sessions.

8. Nonequivalence of groups

Any reason to believe that treatment and control groups might not have been equivalent before treatments? Not reported.

Were steps taken in statistical analyses to adjust for any lack of equivalence? Not reported.

9. Result (for each measure)

Student reading achievement:

a. Name of measure: Stanford Diagnostic Reading Test

A significantly higher number of students of treatment teachers (about 80%) made gains on the reading comprehension subtest than did students of control teachers (about 50%).

There was no significant difference in the number of students of treatment teachers and the number of students of control teachers who made gains on the phonics and the structural analysis subtests.

Teacher effectiveness:

b. Name of measure: Videotaped teaching sessions: Dimensions of teacher shifts.

The treatment teachers showed large significant improvements across all dimensions.

Value of effect size:

1. Treat reading problems openly: +3.8
2. Focus on how to solve problems: +2.80
3. Provide models of thinking: +3.25
4. Teach question-asking: +2.00
5. Ask thought-provoking questions: +3.14
6. Allow student control: +2.08
7. Focus on group collaboration: +2.56
8. Inform students: +2.35
9. Focus on text and learning about reading: +2.52
10. Set reading goals before reading: +3.99
11. Problemsolve during reading: +5.73
12. Summarize to check comprehension: +1.90
13. Reflect on reading goals after reading: +2.21
14. Stress new learning from text: +2.45

Type of summary statistics from which effect size was derived: t-tests

Number of classrooms providing the effect size information: *Ns = 9 and 7*

c. Name of measure: Videotaped teaching sessions: Dimensions of student shifts.

The students of treatment teachers showed large significant improvements across all dimensions.

Value of effect size:

1. S: Focus on how to solve problems: +3.24
2. S: Treat reading problems openly: +3.2
3. S: Express thinking: +2.85
4. S: Ask questions: +2.01
5. S: Give elaborated answers: +1.48
6. S: Take teacher role +2.74
7. S: Focus on group collaboration: +2.46

8. S: Involvement in session: +2.14

Type of summary statistics from which effect size was derived: t-tests

Number of classrooms providing the effect size information: Ns = 9 and 7

Name of measure: Videotaped teaching sessions: Teaching incidents involving problem-solving and collaboration

Treatment teachers showed a far greater percentage of teaching incidents that involved problem-solving and collaboration at posttest than at pretest. No statistical test is presented.

Name of measure: Videotaped teaching sessions: Student and teacher talk

There was a significant increase (t-test) in student talk and a decrease in teacher talk in the treatment condition. The relevant data are not presented.

Brown, Pressley, et al. (1996)

1. Reference

Brown, R., Pressley, M., Van Meter, P., & Schuder, J. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. Journal of Educational Psychology, 88(1), 18-37.

2. Research Question

The purpose of this research was to evaluate the effectiveness of the Students Achieving Independent Learning (SAIL) program. Three hypotheses were examined:

Participating in SAIL would enhance reading comprehension as measured by a standardized test.

After a year of SAIL instruction, there would be clear indications of students learning and using strategies.

Students would develop deeper, more personalized, and interpretive understandings of text after a year of SAIL.

3. Sample of Student Participants

States or countries represented: Mid-Atlantic state (unnamed), United States

Number of different schools: Not reported; all schools in the same district.

Number of different classrooms: 10.

Number of participants (total, per group):

SAIL group = 30
Control group = 30
Total = 60

Number per group = 6

The SAIL and non-SAIL reading groups were matched on the basis of school demographic information and the students' fall standardized test performances (see below).

Age: Not reported.

Grade: Second.

Reading levels of participants: Reading below second grade level.

Setting: Not reported.

Pretests administered before treatment:

Comprehension subtest of the SAT. (Primary 1, Form J; Grade level 1.5 to 2.5); administered in late November or early December.

Special characteristics:

SES: Not reported.

Ethnicity: Not reported.

Exceptional learning characteristics: None, other than reading below grade level.

Selection restrictions used to limit the sample of participants:

Only six students in one SAIL class met eligibility requirements so the researchers decided to use six matched pairs in each classroom as the basis of comparison.

Contextual information (concurrent reading instruction that participants received in their classrooms during the study): Not reported.

Description of curriculum/instructional approach:

413

The SAIL program uses a TSI approach to teaching reading comprehension to low-performing students. The TSI approach contains all the elements of DE and also includes extended discussions that emphasize joint construction of text interpretations and student strategy usage.

"The purpose of SAIL is the development of independent, self-regulated meaning-making from text."

Students are taught to adjust their reading to their specific purpose and to text characteristics.

According to the authors, "the short-term goal of TSI is the joint construction of reasonable interpretations by group members as they apply strategies to texts. The long-term goal is the internalization and consistently adaptive use of strategic processing whenever students encounter demanding text. Both goals are promoted by teaching reading group members to construct text meaning by emulating expert readers' use of comprehension strategies."

SAIL teachers are taught to achieve the goals of TSI through:

- Direct explanations
- Modeling
- Coaching
- Scaffolded practice.

In addition, SAIL teachers are taught to facilitate extended discussions of text, which emphasize student application of strategies to text comprehension.

In the SAIL reading program, students are taught strategies for adjusting their reading to their specific purpose and to text characteristics. Specifically, students are instructed to:

- Predict upcoming events
- Alter expectations as text unfolds
- Generate questions and interpretations while reading
- Visualize represented ideas
- Summarize periodically
- Attend selectively to the most important information
- Think aloud as they practice applying comprehension strategies during reading instruction.

Overreliance on any one strategy is discouraged. In general, students are taught that getting the overall meaning of text is more important than understanding every word.

When SAIL instruction occurs in reading groups, it differs in a number of ways from more conventional reading group instruction:

Prereading discussion of vocabulary is eliminated in favor of discussion of vocabulary in the context of reading.

The almost universal classroom practice of asking comprehension-check questions as students read in group (e.g., Mehan, 1979) is rarely observed in transactional strategies instruction groups (Gaskins et al., 1993). Instead, a teacher gauges literal comprehension as students think aloud after reading a text segment.

There are extended interpretive discussions of text, with these discussions emphasizing student application of strategies to text.

Although reading group is an important SAIL component, the teaching of strategies extends across the school day, during whole-class instruction, and as teachers interact individually with their students. Reading instruction is also an across-the-curriculum activity.

How was sample obtained?

The five SAIL teachers exhausted the pool of 2nd grade teachers in the district with extensive experience (i.e., 3 or more years) teaching in the SAIL program. The comparison teachers were recommended by principals and district reading specialists.

Attrition

Between the first and second semesters, one SAIL student and two comparison students in one pair of classrooms left their classrooms. Backup students were substituted, with no significant difference occurring between the newly constituted groups on the fall reading comprehension subtest.

4. Setting of the Study

Elementary school classrooms.

5. Design of the Study

Quasi-experimental, in that teachers and students were not randomly assigned to conditions.

The authors state that "preparing teachers to become competent transactional strategies instructors is a long process; therefore, the Panel felt that it could not randomly assign teachers, provide professional development, and wait for teachers to become experienced in teaching SAIL in a realistic time frame."

However, as noted above, each of the SAIL groups was matched with a comparison group that was "close in reading achievement level at the beginning of the study" (based on standardized test performance) and from a school that was "demographically similar to the school the school representing the SAIL group."

6. Independent Variables

a. Treatment variables

Describe all treatments and control conditions:

The treatment (SAIL) teachers were not trained specifically for this study; however, they all had extensive experience (i.e., 3 or more years) teaching in the SAIL program.

The control teachers received no special training; however, they were all "highly regarded for their teaching abilities by district personnel." In addition, the control group had, on average, a greater number of years of teaching experience than the treatment teachers.

Was instruction explicit or implicit? Explicit.

Difficulty and nature of texts used:

It is not entirely clear what texts were used during the course of the school year. The three texts used in the study for group comparisons were illustrated stories from trade books, with numbers of words and readability levels as follows:

- 341 words; 2.4
- 512 words; 2.2
- 129 words; 3.9 (used for a different measure than the previous two).

Was trainers' (teachers') fidelity in delivering treatment checked?

The article states that there were "informal observations of the comparison teachers over the year, [which] confirmed that they were more eclectic in their approach to reading instruction than the SAIL teachers . . . ." However, it does not indicate whether the SAIL teachers were also observed.

Properties of trainers (teachers):

Number of trainers (teachers):

SAIL group = 5

Control group = 5

Total = 10

Teacher student ratio:

It is unclear how many students were in each teacher's class; however, the reading groups within each class that were compared had six students each, for a ratio of 1:6.

Type of teacher: Classroom teacher.

Any special qualification of trainers (teachers)?

All the SAIL teachers had between 3 and 6 years of experience teaching in the SAIL program; therefore, one may assume that they delivered the treatment effectively.

The SAIL teachers had an average of 10.4 years of teaching experience compared to an average of 23.4 years for the comparison teachers.

The authors acknowledge that given this difference, "there is no way to separate out the effects that years of experience may have had on the way teachers taught their students."

However, they state that readers should "bear in mind that the comparison teachers were highly regarded for their teaching abilities by district personnel; therefore, if anything, their greater number of years of experience could be construed as an advantage."

Length of training given to trainers: Not reported.

Source of training: Not reported.

Assignment of trainers (teachers) to groups:

"The five SAIL teachers exhausted the pool of 2nd grade teachers in the district with extensive experience teaching in the SAIL program."

Cost factors: Not reported.

Moderator variables:

List and describe other nontreatment independent variables included in the analyses of effects: None reported.

7. Dependent (Outcome) Variables:

Student reading achievement:

SAT:

The comprehension and word skills subtests were used.

(PRE and POST).

Story recall:

Students were asked cued and picture-cued retelling questions about two stories. This measure was designed to assess both recall skills and the degree to which students were interpretive in their retelling of the story. (POST).

Student strategy awareness:

Strategy awareness interview:

In October and November (i.e., when SAIL components were being introduced to SAIL students) and in March and April, a strategies interview was administered to all students participating in the study. This interview tapped students' reported awareness of strategies, as measured by the number and types of strategies they claimed to use during reading. It was also designed to measure students' awareness of where, when, and why to use strategies.

(DURING).

Students were asked the following five open-ended questions, adapted from the ones used by Duffy et al. (1987):

What do good readers do? What makes someone a good reader?

What things do you do before you start to read a story?

What do you think about before you start to read a story?

What do you do when you come to a word you do not know?

What do you do when you read something that does not make sense?

Student strategy usage:

Think-aloud measure:

Students were stopped at four points while reading a difficult story individually with a researcher and asked to describe their thinking and their strategy usage.

(POST).

Teacher effectiveness:

Classroom observation:

SAIL and non-SAIL teachers were observed teaching two story lessons and were compared in terms of the number of strategies they taught in each lesson.

(DURING).

8. Nonequivalence of groups

Any reason to believe that treatment and control groups might not have been equivalent before treatments?

Although it is possible, because the groups were not randomly assigned, it is unlikely because of the careful matching done in the fall on both mean performance and variability on standardized reading comprehension tests.

Were steps taken in statistical analyses to adjust for any lack of equivalence?

Not reported.

9. Result (for each measure):

Student reading achievement:

a. Name of measure: SAT: Comprehension

Students of treatment teachers scored significantly higher than students of control teachers on the comprehension subtest of the SAT.

Value of effect size: + 1.70

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

b. Name of measure: SAT: Word skills

Students of treatment teachers scored significantly higher than students of control teachers on the word skills subtest of the SAT.

Value of effect size: +1.67

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

c. Name of measure: Story recall: Literal information

Students of the treatment (SAIL) teachers recalled more literal information than students of control teachers.

Value of effect size:

Story 1: +0.69

Story 2: +1.37

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

d. Name of measure: Story recall: Interpretation

Students of the treatment (SAIL) teachers were significantly more interpretative in their retelling of the stories than were students of control teachers.

Value of effect size:

Story 1: +1.01

Story 2: +1.07

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

Student strategy awareness:

e. Name of measure: Strategy awareness interview: Comprehension strategies

Toward the end of the treatment, the students of the treatment (SAIL) teachers reported more awareness of comprehension strategies during the interview than did the students of control group teachers.

Value of effect size: +4.03

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

f. Name of measure: Strategy awareness interview: Word-level strategies

Toward the end of the treatment, the students of the treatment (SAIL) teachers reported more awareness of word-level strategies during the interview than did the students of control group teachers.

Value of effect size: +1.38

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

Student strategy usage:

g. Name of measure: Think-aloud measure

Students of treatment (SAIL) teachers applied significantly more strategies during the think-aloud task than did the students of control teachers.

Value of effect size: +2.98

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

Teacher effectiveness:

h. *Name of measure: Classroom observation: Comprehension strategies*

The treatment (SAIL) teachers were found to have taught significantly more comprehension strategies than control teachers.

Value of effect size: +5.48

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5

i. *Name of measure: Classroom observation: Word-level strategies*

The treatment (SAIL) teachers were found to have taught significantly more word-level strategies than control teachers.

Value of effect size: +1.38

Type of summary statistics from which effect size was derived: t-test

Number of classrooms providing the effect size information:

Ns = 5 and 5.

CHAPTER 5

Teacher Education and Reading Instruction

TEACHER EDUCATION AND READING INSTRUCTION

Executive Summary

Introduction

The analysis of reading and reading instruction involves four interacting factors: students, tasks, materials, and teachers. It has often been the case that research has not focused on teachers; it has emphasized students, materials, and tasks. Recent developments, such as class-size reduction and the development of standards for content areas, have highlighted the need for qualified teachers. In addition, teacher education and professional development emerged as one of the most frequently mentioned areas of concern during the regional meetings. Speakers at meetings of the National Reading Panel (NRP) also emphasized the need for consideration of these topics. Given these concerns, a subgroup was established to survey the research in this area. The following is a summary of that work.

Background

Teacher education and professional development represent two aspects of the ways in which teachers acquire knowledge. In teacher education programs, prospective teachers are taught in structured programs before being certified as teachers. The experiences these preservice teachers have include coursework in theory and methods as well as supervised teaching. Once teachers are in the field, having assumed teaching positions, the emphasis shifts from teacher education to professional development. This latter context is often referred to as inservice education. Because there are dramatic differences in the amount of time spent, the structure of the program, and the continuity of the education, the NRP has chosen to analyze the two contexts separately.

The analysis was guided by three primary questions:

- How are teachers taught to teach reading?
- What do studies show about the effectiveness of teacher education?
- How can research be applied to improve teacher development?

Two secondary questions were posed before the analysis:

- What findings can be used immediately?
- What important gaps remain in our knowledge?

Methodology

How was the analysis of the research literature conducted?

The NRP conducted extensive and systematic searches for research on preservice and inservice teacher education and professional development. According to the methodology developed by the NRP, only studies that were experimental tests of teacher education or professional development and that had appeared in professional journals were included. Each study that met the initial criteria was coded with variables that allowed for further analysis.

Results and Discussion

What do the results of the analysis of studies on teacher education and reading show?

Despite the fact that there is a much larger body of work on teacher education, only a very small number of studies were found to meet the initial criteria. There were differences between the types of problems studied in preservice and inservice research. Preservice research emphasized the learning of methods and use of materials. Inservice research was much more eclectic, seemingly related to specific curricular needs rather than the general instructional needs at the preservice level.

A second important issue is whether teacher education is effective. For teacher education to be effective, it must change both teacher and student behavior. That is, teachers must adopt new ways of teaching, and students must show appropriate improvement as a result. However, it is only for inservice research that student achievement was measured. For preservice work, only teacher outcomes were measured. This is not entirely inappropriate because this research does show that teachers adopt the strategies and techniques they are taught.

Of the inservice research studies, one-half measured student outcomes as well as teacher outcomes. In all but a few cases the results showed that the intervention in professional development produced significantly higher student achievement.

Because of the small number of studies that constituted the final sample, the Panel could not answer the question of how research can be used to improve teacher education in specific ways. Rather, it is clear that there is a need for programmatic research to answer this question.

Additional evidence on this issue is available in the report from the Comprehension subgroup. The conclusion with respect to the preparation of teachers for comprehension instruction is that it requires extended training with ongoing support. That only a few studies were found dealing with teacher education and professional development in comprehension supports the conclusion of this analysis that a great deal of research is needed on this issue.

Almost all the research demonstrated positive effects on students, teachers, or both. However, the range of variables was so great for the small number of studies available that the NRP could not reach a general conclusion about the specific content of teacher education programs.

Conclusions

What conclusions can be drawn from this analysis of teacher education and studies?

Based on the analysis, the NRP concludes that appropriate teacher education does produce higher achievement in students. Much more must be known about the conditions under which this conclusion holds. Some issues that need to be resolved include determining the optimal combination of preservice and inservice experience, effects of preservice experience on inservice performance, appropriate length of interventions for both preservice and inservice education, and best ways to assess the effectiveness of teacher education and professional development.

Directions for Further Research

There was little research on how teachers can be supported over the long term to ensure sustained implementation of new methods and student achievement. This is an important issue that needs resolution, given the resource-intensive nature of teacher education and professional development.

The Panel found no research in the sample that addresses the question of the relationship between the development of standards and teacher education or professional development. Given the great interest in developing standards, this is an important gap in our knowledge.

TEACHER EDUCATION AND READING INSTRUCTION

Report

Introduction

The analysis of reading and reading instruction involves four interacting factors: students, tasks, materials, and teachers. It has often been the case that research has not focused on teachers, emphasizing students, materials and tasks. Recent developments such as class-size reduction and the development of standards for reading and content areas have highlighted the need for, and difficulty in obtaining, qualified teachers. Although accreditation processes for schools and colleges of education (National Council for Accreditation of Teacher Education, for example) and certification of programs (Association for Childhood Education International and International Reading Association) exercise some control over the quality of teacher preparation, there is a need for the standards utilized by these governing bodies to be validated by and predicated on empirical research. (Versions of standards presently used for accreditation related to reading literacy are found in Appendix C.)

Teacher education and professional development emerged as being among the most frequently mentioned areas of concern during the regional meetings. Speakers at meetings of the National Reading Panel (NRP) also emphasized the need for consideration of these topics. Given these concerns, the NRP agreed to include a survey of the research in this area in its report.

Gordon (1985) believed that teacher education originally (19th century origins) and to date was and is largely designed as vocational training, based on an apprenticeship model of education lending its programs to behavioristic learning, imitation, and repeated practice. In addition, it has been almost an article of faith among many teacher educators that there is a body of knowledge that can (and should) be learned as a major component of learning to be a teacher. (See, for example, Shulman, 1986). In addition, Shulman (1986) called for teacher education to be "research-based." Whereas most proposals for improving teacher education have presumed to draw on the research literature, those proposals have not unequivocally called for the research-based evaluation of teacher education itself.

There is a growing body of research that shows correlations between aspects of formal teacher preparation and quality of teaching or student outcomes. In a recent study, Darling-Hammond (2000) showed that teacher quality characteristics such as certification status and degree in the field to be taught are significantly and positively correlated with student outcomes. Darling-Hammond (2000) also reports that "NAEP [National Assessment of Educational Progress] analyses found that teachers who had had more professional training were more likely to use teaching practices that are associated with higher reading achievement on the NAEP tests."

However, there are important caveats associated with this work. It is correlational and, although suggestive, does not deal with the detail necessary to provide specific recommendations for teaching. There is no way to determine what variables account for the general relationship. Research that demonstrates causal relationships might provide more consistent guidance. Moreover, the work does not give much guidance about what the content of teacher education or professional development programs should be.

Other types of reading intervention have also emphasized teacher education in a variety of ways. Notable among these is Reading Recovery®. Jongsma (1990) suggests that teachers go through a type of "retraining" because Reading Recovery® introduces new ways of looking at literacy learning. By implication, all new ways of looking at reading would require some professional development. Clay (1991) points out the importance of the initial "training" and subsequent needs for inservice development.

A note on usage is appropriate here. The NRP has chosen to use the phrase teacher education rather than teacher training to reflect what the Panel believes is the professionalization of teachers and teaching. Although it is possible to "train" teachers to use particular methods to teach, it seems more appropriate to educate teachers in a professional context that will give them control over a wide range of decisionmaking tools.

The Panel also distinguishes between teacher education (largely preservice or prior to certification) and professional development (largely inservice or postcertification). The Panel has done this for two reasons. First, it is conceptually important to distinguish between programs in which participants are essentially full-time students and part-time teachers and those in which participants are full-time teachers and part-time students. The second reason is that the research fell into these distinct categories. Different concerns and different research variables and outcomes were involved in the two different research literatures. Despite the division, the Panel does believe they are clearly related.

Taken together, the many theoretical formulations, empirical findings, and practical concerns suggest how important teacher education is in the teaching of reading. It was deemed appropriate to conduct an analysis of the research on teacher education to determine what can be supported by research.

The analysis was guided by the three primary questions:

1. How are teachers taught to teach reading?
2. What do studies show about the effectiveness of teacher education?
3. How can research be applied to improve teacher development?

Two secondary questions were also posed prior to the analysis:

1. What findings can be used immediately?
2. What important gaps remain in our knowledge?

Methodology

There is a widespread belief that there is little research on teacher education, despite the great interest in the issue. Cruickshank and Metcalf express this sentiment:

> Literature on the conduct, objectives, and the effectiveness of training in teacher education is sparse . . . . Given the historic brouhaha over training in teacher preparation, it would be expected that a considerable available related literature would exist. Such is not the case (Cruickshank & Metcalf, 1990, p. 491).

Database

To examine the research related to teacher education and professional development, electronic searches were performed on the ERIC, PsycINFO, OCLC World Catalog, and OCLC Article First databases. The search terms used and numbers of articles returned are included in Appendix A.

The initial selection process identified more than 300 papers; many of these were nonexperimental and were therefore not included. The resultant set of studies was then divided into two categories: research on preservice and research on inservice or professional development. The criteria used were that preservice research was primarily concerned with the training of prospective teachers before certification or full-time work in classrooms, whereas inservice work was concerned with teachers who were already teaching in school environments.

To supplement the electronic searches, the bibliographies of the articles identified in the electronic searches and a recent review of teacher education research in reading (Anders, Hoffmann, & Duffy, 2000) were examined for additional citations that did not appear in the electronic searches themselves. Appropriate citations that had not been identified in the electronic searches were added to the pool of research studies to be examined. There were four studies reviewed in the comprehension subgroup report on preparing teachers to teach reading comprehension. Those four studies were included in the teacher education analysis as well.

A total of 32 studies met the final criteria: 11 preservice and 21 inservice. Because of the way in which the results of some of the underlying research was reported, there were more articles than studies. That is, there were two instances where two published papers reported on different aspects of the same research project. An additional eight studies focused on inservice on teaching for special education or learning disability students. These have not been coded but are noted here as a subgroup of the inservice studies.

Analysis

It was determined that to conduct meta-analyses on these data would be inappropriate because there is not a critical mass of studies researching the same variables or theoretical positions. Moreover, although all the studies do address the general problems of improving teacher education, the underlying rationales for the studies represent an eclectic mix of theories and conceptualizations.

Consistency With the Methodology of the National Reading Panel

The methods of the NRP were followed in the conduct of the literature searches and the examination and coding of the articles obtained. Because a meta-analysis was deemed inappropriate, the data were coded using a subset of the coding scheme adopted by the NRP. These data are contained in Appendix B.

Some Additional Considerations in Research on Teacher Education

When research is conducted on instructional variables, it is often the case that the participating teachers receive instruction in the instructional interventions. For example, when comprehension strategy research is conducted in classrooms, the instructors (either classroom teachers or the researchers) must be taught to conduct instruction in the appropriate manner. In this sense, almost all of the research the NRP has identified contains some elements relative to teacher education. However, in these circumstances, the focus is almost exclusively on student outcomes, without detailed data on changes in teacher behaviors. Although the NRP recognizes the importance of the more general form of teacher education and professional development, it determined that these factors would not be included in the current analysis because of the lack of teacher performance data.

There are also notable programs where teacher education or professional development is an important component of the intervention. Reading Recovery© is one example of such a program; Success for All is another. However, most of the research studies on these programs do not include measures of teacher changes in their results. Again, as in most instructional research, the focus is on the specific interventions and student outcomes rather than teacher change. The Panel did not find studies that met the NRP criteria that were in either of the two categories.

One reason that teacher education has been ignored in these research contexts is that researchers believe that any changes in student outcomes are attributable to the intervention, which is, in turn, delivered by the participating teachers. This would logically imply that teachers had learned to deliver the instruction in the way the research program dictated. This is, in part, the criterion of fidelity to the intervention. However, the issue goes well beyond fidelity of teaching to the many other variables that relate to teaching rather than to learning.

Although these studies have *not* been analyzed as part of the pool of studies, they have some relevance to the interpretation of the analysis. Consequently, recommendations at the end of the analysis have been influenced by these concerns.

Results

In the presentation of results, the research on preservice teacher education has been separated from that on professional development with inservice teachers. The Panel believes this is fundamentally appropriate because different quality criteria and outcome measures can be applied to the research studies. In particular, the criteria of success are different for the two sets of studies.

That is, for preservice studies, the focus is almost entirely on changing teacher behavior, without a concomitant focus on the outcomes of students who are (eventually) instructed by those teachers. The Panel found no instances of research in the pool that continued with preservice teachers as they moved into full-time teaching positions. There is no inherent reason why this is the case. The reasons seem, instead, to be pragmatic and related to the complexities of research that would be introduced in attempting to follow teachers into full-time teaching. Although the lack of student data limits the conclusions one can draw about the results of this research, it does provide an important background for other teacher education and professional development research. If teacher behaviors cannot be transformed by changes in the curriculum in preservice programs, it is unlikely that teacher behaviors can be changed later.

For inservice research, the ultimate test of success is whether students benefit from instruction delivered by teachers as a result of that intervention. Consequently, the Panel invoked a strong criterion that student outcomes must be part of the research on inservice teachers. However, another criterion is also critical. If there is no change in teachers as a result of the intervention, it is not possible to attribute changes in student outcomes to the teacher development intervention. Other factors must be invoked to account for the changes in students. Consequently, the NRP must have both teacher changes and student changes to agree that inservice interventions are effective. Although the Panel believes that preservice and inservice research form two different bodies of work, they are related in that preservice does provide evidence for the efficacy of producing teacher change. Those changes can be important in designing inservice interventions.

Preservice Studies

Eleven preservice studies met the criteria for this portion of the NRP analysis. These preservice studies, with coded information, are grouped in Table 1 in Appendix B. Table 2 in Appendix B lists two studies that involved preservice interventions as well as inservice interventions. Most of the preservice research (ten studies) focused on elementary reading instruction. Two (of the ten) studies had a broad range of grade samples, spanning grade levels from K through 8 and 1 through 6. For one study it was not possible to determine the grade level.

The content of the teacher education in these studies is a primary variable in distinguishing among studies. The 11 studies can be classified into the following four categories. For each category, the number of studies is indicated in parentheses.

- **Comprehension and strategy instruction:** Questioning techniques (2)
- **General methods:** Directed Reading-Thinking Activities (DRTA); teaching word recognition skills; Directed Reading Activity (DRA); Informal Reading Inventory (IRI) (4)
- **Materials:** Estimating readability levels; teacher decisionmaking and awareness of materials (2)
- **Others:** Case method; study skills; theoretical orientations to reading (3)

The majority of the preservice studies reviewed (10 of 11) reported improvements in teacher knowledge. Of these ten, two reported mixed or modest effects. Only one study, which looked at the accuracy of teachers in estimating the readability levels of materials, did not report any effect from having either theoretical knowledge of reading or teaching experience, or both, compared with a control group with neither theoretical knowledge nor teaching experience.

The duration of the studies reviewed here ranged from 5 to 6 weeks to about a year, which corresponds closely to the structure of university-based coursework. Although these studies show that preservice courses improved prospective teachers' knowledge, there is no way of knowing whether this increased knowledge actually translates into effective teaching because none of the studies reports data on the teachers after their participation in the experimental program.

In the NRP sample, no studies of larger scale interventions at the program level were found. For example, there were no experimental studies that looked at changes in the format of teacher education programs like the use of professional development sites or the use of standards-based programs.

Inservice Studies

There were 21 inservice studies that met the criteria for this review. These studies are listed in Appendix B: Coding of Studies. There are four groupings: studies that involved both inservice and preservice interventions (Table 2), studies that measured only teacher outcomes (Table 3), studies that measured both teacher and student outcomes (Table 4), and studies that measured only student outcomes (Table 5).

The first analysis of the data was to determine the grade levels of the teachers who participated in the inservice work. For 18 of the studies, it was possible to do so. Because the studies often involved multiple grade levels, there was a total of 70 different samples of teachers represented in the 18 studies. These data are represented in Figure 1 on the next page.

It is evident that the inservice instruction is targeted at the elementary grades with approximately equal emphasis. The numbers of studies across grades 1 through 5 are equal. There are far fewer studies at the middle and high school grade levels, with only a single study at each of the high school grades.

A second analysis examined the focus of inservice instruction for teachers of reading. Compared with the work in preservice programs, inservice instruction seems to be more eclectic, ranging from training in specific methods (e.g., how to use reading groups) to more extensive instruction encompassing ways to teach reading, classroom management, and lesson design. The topics fell into the following categories, with the number of studies indicated in parentheses.

- **Comprehension and strategy instruction:** Higher order questioning, explicit instruction in using reading skills strategically; questioning and student-teacher interactions; Transactional Strategy Instruction (TSI); questioning and response guidance cues (8)
- **General methods:** Skills vs. Language Experience Approach (LEA); DRA; whole language; phonics, question-and-answer, and giving feedback; teaching a language arts integrated curriculum (5)
- **Classroom management:** Small groups; reading groups; conducting cooperative learning activities; using performance assessment; translating Madeline Hunter's Instructional Theory Into Practice, focusing on effective classroom management, motivation and lesson design (5)
- **Improving teachers' attitudes:** Teaching writing as a process to facilitate change in teachers' attitudes to language; improving content area teachers' skills and attitudes to teaching reading; enthusiasm training. (3)

It appears to be the case that the emphasis is on specific methods of teaching reading, rather than the general methods that characterize preservice research. There is much less emphasis on the general aspects of teaching reading. Three studies investigated ways in which to improve teacher attitudes, reflecting the needs of teachers on the job.

Effectiveness of Inservice Instruction

Only 11 studies in the NRP pool measured *both* teacher and student outcomes. Six other studies measured only teacher outcomes, whereas four measured only student outcomes. As noted above, it is necessary to have both teacher and student outcomes to be able to determine whether teacher education is effective. If it is, it must change both teacher and student behavior. That is, teachers must adopt new ways of teaching and students must show appropriate improvement if the results are to be attributed to the new ways of educating teachers.

The measures of teacher change and student outcomes used in this body of research were a combination of informal, researcher-designed assessments and standardized evaluations. As a generalization, the teacher outcome measures were all researcher-designed, whereas the student measures tended to be standardized instruments. At times, student outcomes were measured with a combination of researcher-designed and standardized measures. Given that the researchers designed the treatments, standardized measures of outcomes often did not exist, necessitating the development of researcher-designed instruments.

Another set of analyses examined the duration of the project and the number of hours of instruction delivered. Figure 2 presents the data on the duration of projects.

Of the 21 studies, only 4 had durations of 6 months or less. However, the duration of the project is not necessarily the crucial variable. Where possible, the total amount of time spent in instruction was also examined. It was possible to determine the number of hours of instruction in 11 studies. For many of the studies, the number of hours of instructional intervention is not specified; these studies were not included in this analysis. Often what are reported are phrases like "a monthly meeting" or "weekly workshops." No attempt was made to interpret these; only those studies for which unambiguous determinations could be made were analyzed. The data for instructional time are presented in Figure 3.

Figure 3 shows that for the 12 studies for which instructional time could be determined, the greatest number of hours of instruction was 60. The majority of the studies (8 of 12) presented 15 or fewer hours of instruction.

Figure 1. Number of Studies as a Function of Grade Levels of Teachers for Inservice Research (18 Studies with 70 Grade Samples)

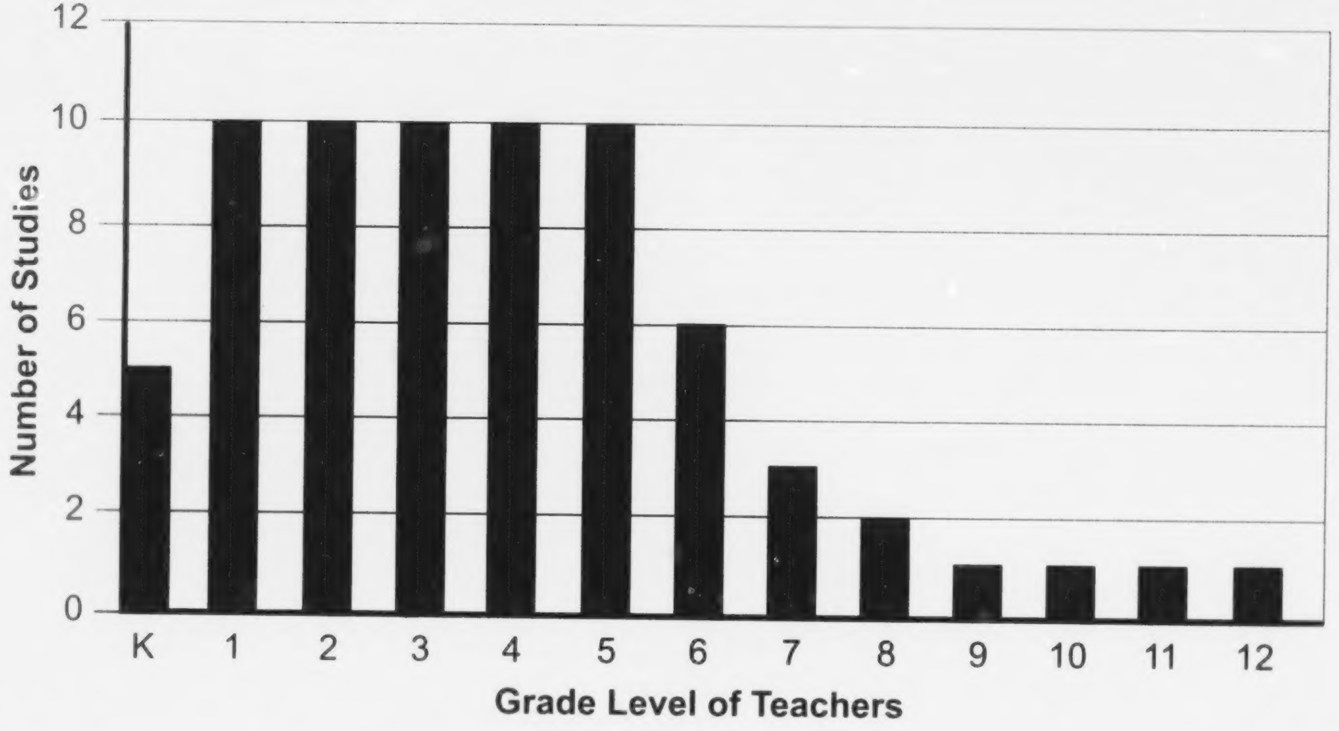


Figure 2. Number of Studies as a Function of Duration of Inservice Projects (N=20)

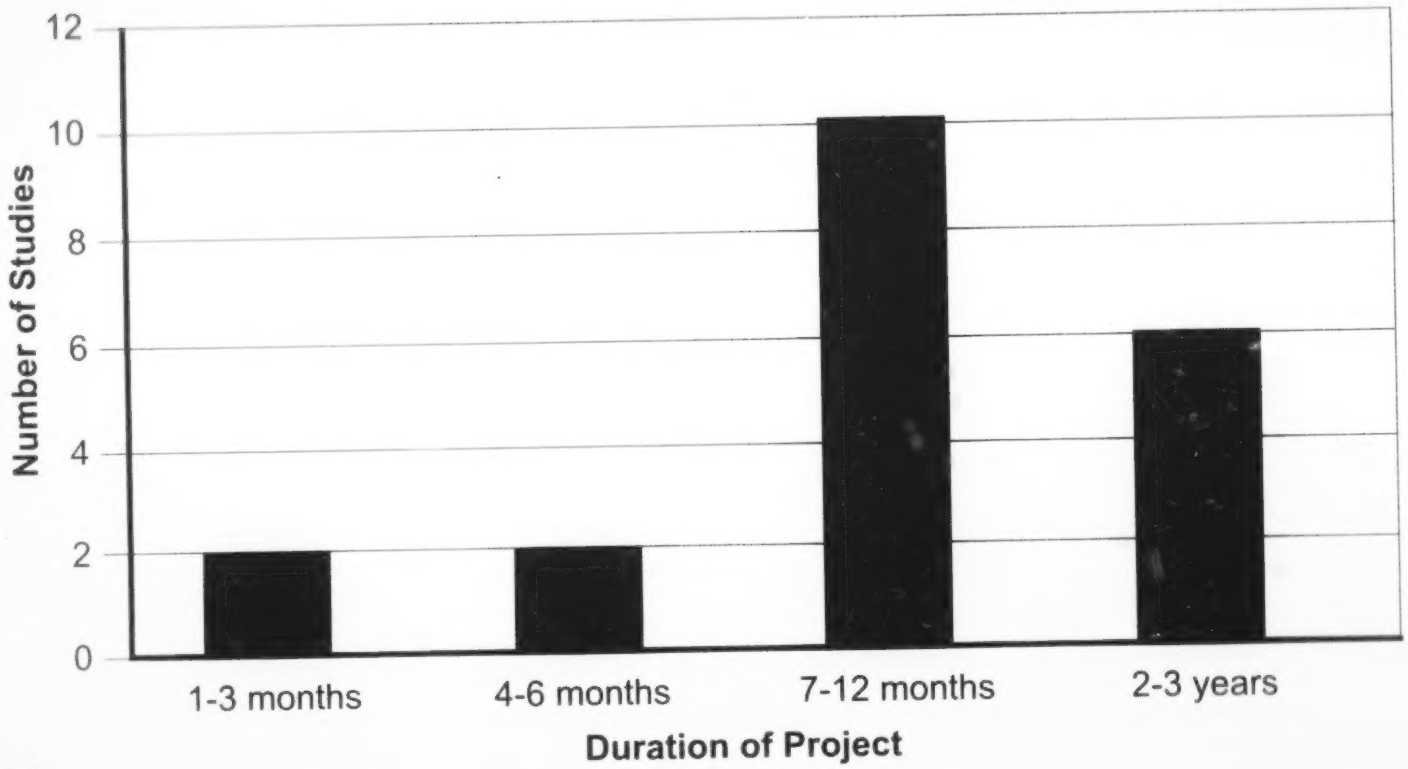
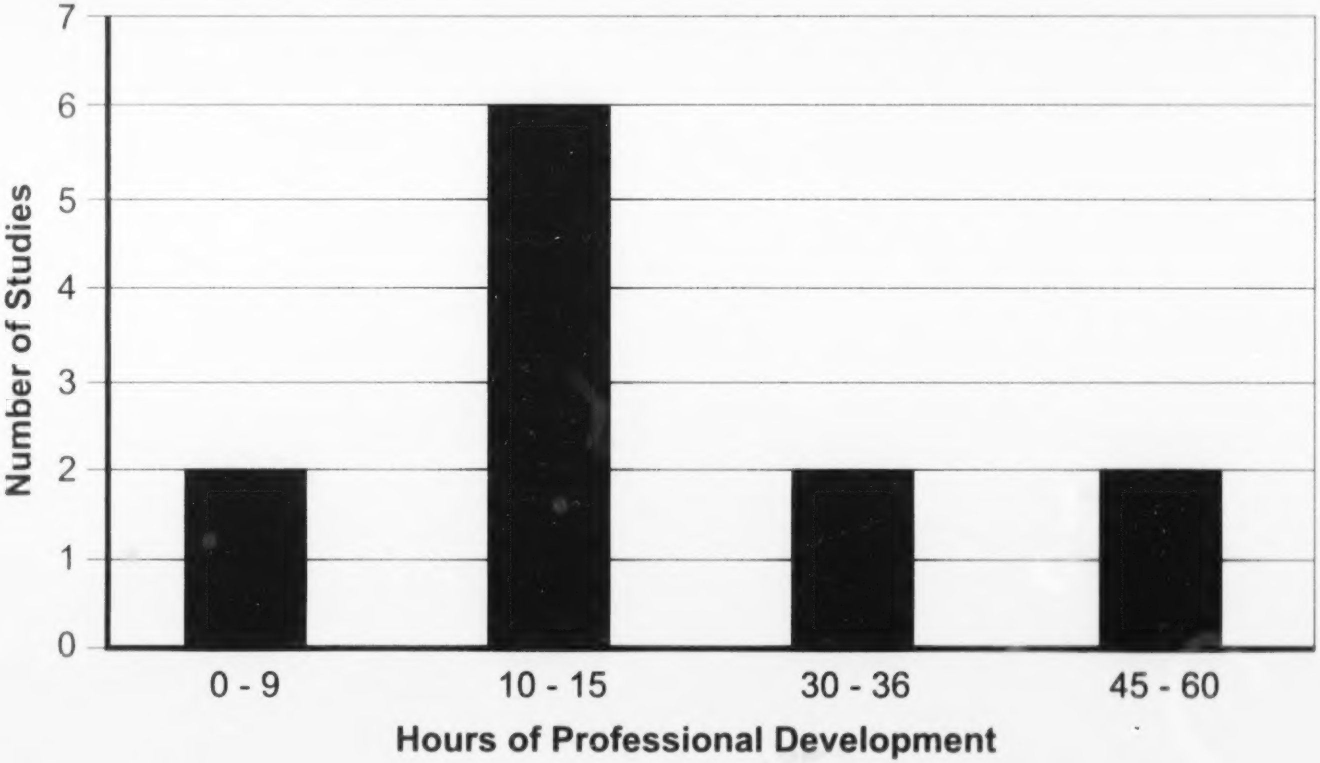


Figure 3. Number of Studies as a Function of Amount of In-Service Professional Development, (N=12)



Studies Reporting Positive Changes in Teacher Outcomes

Seventeen out of the 21 studies reviewed measured teacher outcomes. Fifteen of these studies reported significant or modest improvements in teachers' knowledge or practice. Out of the fifteen studies that measured student outcomes, 13 reported improvements in student achievement. One clear trend in the data is that where teacher outcomes showed significant improvement, so did student achievement. In studies where no gains are reported for the teachers, no gains are reported for the students in the same study. In general, one can conclude that inservice professional development does lead to improved teacher knowledge and practice and improved student achievement. Because the content of each of these studies is widely divergent, it is not possible to reach a specific conclusion about the content of instruction.

Studies Reporting No Change in Teacher Outcomes

Three studies (Coladarci & Gage, 1984; Morrison, Harris, & Auerbach, 1969; Stallings & Krasavage, 1986) reported no change in teacher outcomes, in at least some of the conditions in the research projects. In two of these studies, where student outcomes were measured, student achievement did not improve either.

A closer look at these studies reveals two interesting points. First, one study (Coladarci & Gage, 1984) did not involve *any* formal instruction for teachers. Instead, teachers in the treatment group were given "teacher education packets" consisting of materials on a diverse range of topics, including behavior management, large-group instruction, use of question-and-answer, phonics, questioning, and feedback strategies.

Second, all three studies were long-term projects. The study in which teachers received no formal instruction lasted about a year. The other two were 3 years in duration. Morrison and colleagues (1969) caution against using short-term results to validate teacher education efforts because, in the course of their 3-year study, they found that teachers and administrators reverted to what they had been doing before the project began. Stallings and Krasavage (1986), at the end of their 3-year study, also reported that teacher and student outcome measures actually declined although gains by teachers and students were reported during the first 2 years of the study.

However, three long-term inservice programs reported by Talmage, Pascarella, and Ford (1984), Miller and Ellsworth (1985), and Duffy and coworkers (1987a) showed gains by teachers and significant or partial achievement gains by students. Because of this discrepancy, the Panel could find no relationship between the amount of instructional time (or duration of programs) and student outcomes. This may be a function of the limited number of research studies for which the Panel could make the relevant determinations.

It is difficult to compare the studies reviewed here in terms of the duration of instruction that the teachers received. Hence, it is not possible to draw specific conclusions about the relationship between length or intensity of instruction and outcomes. The duration of the inservice intervention depends on the specific objectives and requirements of the program. Sometimes the intervention consisted of the dissemination by mail of a manual (Coladarci & Gage, 1984) or two meetings and the discussion of a teaching manual (Anderson, Evertson, & Brophy, 1979). It could take the form of a series of workshops or meetings spread over 2 days (Scheffler, Richmond, & Kazelskis, 1993) or a year (Shepard, Flexer, Hiebert, Marion, Mayfield, & Weston, 1996) or three workshops spread over 3 summers (Spanjer and Layne, 1983). It could also take the form of a systematic 3-year staff development program (Stallings, Robbins, Presbrey, & Scott, 1986; Stallings & Krasavage, 1986). The studies do not report the duration of the intervention in a consistent manner: some report the number of hours of instruction, whereas others report the overall duration of the project or duration of the staff development program.

Two other issues were difficult to assess. The Panel was unable to determine the amount of resources (personnel, equipment, and materials) from the reports of the research. This amount would have a direct bearing on the ultimate effectiveness of the interventions. It was also not possible to find any experimental research on inservice professional development that related to the issues surrounding standards-based education.

The NRP did not conduct a separate analysis of the research on preparation of teachers for comprehension instruction. An extensive analysis of this research is included in the report from the comprehension subgroup.

Results: Vocabulary Instruction Methods

Summary of Findings

The NRP is encouraged by the fact that there is a growing body of experimental research on teacher education and professional development. Although this body of research does not, at present, converge on highly explicit and specific recommendations for teacher education, it does suggest that teacher education is successful in most contexts. It also clearly indicates that when teacher education is successful, student performance improves as well.

At the outset of the review, five questions were listed that guided this analysis. In the following summary, there are first some general comments about what was found with regard to each of the questions. Following that is a more interpretive summary.

Summary Answers to the Specific Questions for the Review

Unfortunately, the Panel was unable to answer all five questions with the same level of confidence, simply because the data were insufficient. The following paragraphs summarize the information from the analysis relevant to each of the questions.

- **How are teachers taught to teach reading?**

The Panel found no single method that produced results that clearly indicated unquestioned superiority. Rather, an eclectic mix of methods was found that ranged from macro to micro in their focus. There was an emphasis on methods at the preservice levels contrasted with an emphasis on particular instructional problems at the inservice level. As indicated above, there were simply too many approaches in this small sample to allow conclusions about any one specific method.

- **What do studies show about effectiveness of teacher education?**

The set of results for these studies shows overwhelmingly that interventions in teacher education and professional development are successful. That is, teachers can learn to improve their teaching in ways that have direct effects on their students. Although this was demonstrated only for inservice interventions, there is no reason to believe this is not the case for preservice teachers. There is simply no research that demonstrates this in a positive fashion. Because most of the research demonstrates the effectiveness of teacher education interventions, there is no reason to envisage a different outcome for preservice teachers.

Implications for Reading Instruction

How can research be applied to improve teacher development?

Although there is no single, consistent set of findings that points to specific conclusions, the research has some general implications for effective teacher education and development. First, research can determine which of the interventions in teacher education are most effective. Moreover, characteristics of successful teacher education interventions are beginning to emerge. This research suggests that there is a need, particularly at the inservice level, for extensive support (both money and time) on a continuing basis for teacher education efforts. It is also the case that the support must be continued for an extended period of time. The report on Teacher Preparation by the comprehension subgroup reaches similar conclusions.

What findings can be used immediately?

The studies analyzed in this report do not converge on specific findings with regard to content. Rather, the research suggests that teachers can and do learn to change and improve their teaching. So long as the interventions themselves are based on solid research findings, the interventions in teacher education should produce positive results for teachers and for their students. The research does have implications for the manner in which teacher education is conducted. These implications are discussed more thoroughly in subsequent sections.

Additional Conclusions About Teacher Education and Professional Development

The most obvious conclusion about the research reviewed is that it clearly demonstrates that teachers can be taught, in both preservice and inservice contexts, to improve their teaching. For preservice teachers, this means that prospective teachers do adopt the teaching methods and attitudes they acquire during the course of their education. Inservice teachers not only demonstrate improvement in their teaching; this improvement leads directly to higher achievement on the part of their students. These findings were demonstrated in an overwhelming majority of the research studies reviewed.

However, there is insufficient research to draw exact conclusions about the content of teacher education and professional development programs. Rather, a wide range of techniques and content seemed to produce improvement in teaching and in student outcomes. The body of research on these topics is fragmented when it comes to this level of questioning. There are studies of specific methods of teacher education with specific content as well as more general studies that offer no guidance on content.

Teacher attitudes do change as a result of intervention in both preservice and inservice contexts. This is an important finding because it is the predisposition of teachers to change that makes change possible. Without a change in attitude, it is extremely difficult to effect changes in practice. Most of the research that measured attitudes demonstrated that attitudes did change as a result of the interventions, indicating that at least one of the major prerequisites for teacher change can be taught.

Teacher practices improve as a result of education, but it is not clear for how long these changes are sustained. Teachers may use the new methods only when observed. Although some of the studies in this sample were long term, exceeding 2 years, there is little evidence on the sustainability of the interventions. That is not to say that the interventions were not sustained, but that in most of the studies there was simply no evidence presented that spoke to this issue.

Student achievement outcomes can be improved as a result of teacher development. For inservice studies that measured both teacher and student outcomes, this was a clear finding. These studies represent the most effective types of research, recognizing the need to assess both teachers and students. However, even in these studies, sustainability of the student improvements is an issue that was not addressed.

Directions for Further Research

What important gaps remain in our knowledge?

Perhaps the most apparent feature of the research analyzed in this study is that there are significant gaps in our knowledge of teacher education and development across the board. Part of the difficulty is that high-quality teacher education research is expensive and requires intensive collaborative efforts from all the stakeholders. In subsequent sections, the Panel details what it considers the most important questions that need to be resolved.

The Panel found no studies in the sample that addressed questions related to the development of standards. Therefore, it makes no conclusion about the efficacy of establishing either content standards for students or for teaching teachers on the basis of those standards. Many of the interventions clearly include elements that are also contained in many standards-based programs. However, too many other factors are involved to be able to attribute causal relationships.

The Panel also found that the reporting of studies was inconsistent. Many studies were not described in sufficient detail to make comparisons. Foremost was a lack of consistent attention to the amount of instruction and the frequency of instruction in the description of the studies, which makes it difficult to tell whether it was reasonable to expect either success or failure in individual studies. Some studies reported only the number of sessions, others only the amount of instruction, and still others neither.

Another important oversight was a description of the resources (personnel, time, money, facilities, etc.) required to implement particular programs. It was often impossible to tell what it would take to implement some of the interventions. Consequently, no assessment could be made about the cost-effectiveness of most of the programs or interventions.

There is a large body of nonexperimental literature that addresses teacher education issues. Under the guidelines established for the review, this literature was to be used to help interpret findings from the analysis of the experimental literature. However, because of the lack of convergence in the experimental research, the Panel was unable to bring this nonexperimental literature to bear on the current analysis.

The NRP believes that the nonexperimental literature is a rich source for future research programs. Teacher education research involves particularly complex problems. Doing the research is expensive and time consuming. Therefore, one particular contribution of the nonexperimental literature may be to provide a source of problems to be studied under more controlled conditions. That is, the descriptive literature could be brought to bear to reveal current practices, variables, and so forth, that seem promising (or not) under general conditions. Such insights could guide research that looks more closely at causal relationships or in more specific situations. In addition, the Panel refers the reader to the conclusions of the Text Comprehension report, in the belief that the principles underlying them apply more broadly to other subject areas and could also serve to guide future research in teacher preparation.

The small set of experimental studies reviewed does not allow us to address all the questions that originally guided the analysis. Some of these remain unanswered because of the eclectic nature of the work found. Many are unanswered because they were not addressed specifically in the experimental body of research. There was a great deal of nonexperimental research that fell outside the scope of the experimental domain examined. This research addresses a few of the relevant questions that are listed below, but not all and certainly not definitively. A general conclusion here is that although we have a great deal of knowledge about teacher education, much more remains to be learned.

Many of the questions are unanswered because of the resource intensity of teacher education research. It takes a great deal of time and money to do teacher education research in ways that will yield appropriate answers. It takes a commitment from stakeholders, and it takes a great deal of coordination among them. Rarely do all of these elements come together in a way that admits of experimental research.

However, simply providing money and time is insufficient. High-quality teacher education research must bring together persons who are engaged in quite different endeavors in school contexts. They are used to having control over their own domains and often do not want to relinquish control to any outside influences. Moreover, new "alliances" need to be formed. For example, to answer the questions about effectiveness of preservice education, graduating teachers will need to be followed as they assume teaching jobs. Those who do the preparation of teachers will have to work with persons in the new locations where the graduates work. (Because schools rarely hire teachers en masse, the alliances may have to span districts or other geographic locations to be able to study teachers in sufficient numbers.)

To accomplish the kind of reforms that accompany teacher education improvement requires years of sustained effort at keeping all elements of the system in balance. All of this must take place against a backdrop where the participating individuals may change over the course of a research project. Placed against the other demands (tenure, teaching, publication) on many academic researchers, commitments to the long-term nature of teacher education research often seem daunting.

In addition to the appropriate resources, stronger and more coherent conceptualizations of teacher education and professional development are needed. These conceptualizations need to combine research from a wide variety of perspectives and paradigms to provide the most coherent description of teacher education possible. Such conceptualizations will guide research in more systematic ways, rather than allowing the highly eclectic forms of investigations that characterize current teacher education research. There are excellent examples of good teacher education research; more are needed, as is better reporting of the results as they are disseminated so that subsequent research can build on completed research rather than begin anew with each effort.

We need to find out how teachers can be supported over the long term to ensure sustained implementation of new methods or programs, as well as the sustainability of student achievement. There is a trend in the research analyzed that suggests that teachers may revert to their original methods of teaching; it is important to determine how best to have teachers maintain any improvements they make in their teaching abilities.

Another problem that needs to be addressed in teacher education research is the precise nature of the interventions. In the literature the NRP analyzed, there is only sparse information on the precise content of what teachers were taught. Rather, there is a mix of techniques, methods, theories, and materials that are often confounded with each other in the instructional contexts. Some of the instructional methods focus on teacher attitudes while others focus on the use of specific materials. This question should be addressed in a systematic way.

There is a need to develop and refine the ways in which we study the link between teacher education and student outcomes. Only a few inservice studies looked at both teacher and student outcomes. None of the preservice research made the link between teacher outcomes and ultimate student performance. Although all the inservice research that reported improved teacher outcomes also reported improvement in student achievement, there is no evidence that this is true for preservice programs.

Because teacher education is a labor-intensive endeavor, new ways of instruction need to be developed that make it possible for instruction to be more effective. In the sample of studies, the Panel found a total of seven preservice and inservice research studies that used various forms of technology to improve teacher education. This is a promising direction. Computer technology has made the use of video modeling and simulation even more available than it has been. The use of either simulated or real teaching cases, linked with appropriate instruction, can provide supplemental experiences to classroom instruction in teaching.

The list of questions that remains is a long one. However, there is a growing consensus on many elements of the problems in teacher education and professional development. The technology to improve teacher knowledge and performance exists. Positive changes in teacher education have been demonstrated by a wide variety of interventions. Further studies are needed to address the problems that remain.

References

Anders, P., Hoffmann, J., & Duffy, G. (2000). Teaching teachers to teach reading: Paradigm shifts, persistent problems, and challenges. In M. Kamil, P. Mosenthal, P.D. Pearson, & R. Barr (Eds.). Handbook of Reading Research, Vol. 3. (pp. 721-744). Mahwah, NJ: Lawrence Erlbaum Associates.

Clay, M. M. (1991). Why is an inservice programme for Reading Recovery teachers necessary? Reading Horizons, 31(5), 355-372.

Cruickshank, D. R., & Metcalf, K. K. (1990). Training within teacher preparation. In W. R. Houston (Ed.), Handbook of Research on Teacher Education (pp. 469-497). NY: Macmillan Publishing Company.

Gordon, B. (1985). Teaching teachers: "Nation at risk" and the issue of knowledge in teacher education. Urban Review, 17, 33-46.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Education Policy Analysis Archives [On-line], 8(1). Available: http://epaa.asu.edu/epaa/v8n1/

Jongsma, K. S. (1990). Training for Reading Recovery teachers (questions and answers). Reading Teacher, 44(3), 272-275.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4-14.

Appendix A

Studies Analyzed

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. Elementary School Journal, 79(4), 193-223.

Baker, J. E. (1977). Application of the in-service training/classroom consultation model to reading instruction. Ontario Psychologist, 9(4), 57-62.

Block, C. C. (1993). Strategy instruction in a literature-based reading program. Elementary School Journal, 94(2), 139-51.

Book, C. L., Duffy, G. G., Roehler, L. R., Meloth, M. S., & Vavrus, L. G. (1985). A study of the relationship between teacher explanation and student metacognitive awareness during reading instruction. Communication Education, 34(1), 29-36.

Brown, R., El-Dinary, P.B., Pressley, M., & Coy-Ogan, L. (1995). A transactional strategies approach to reading instruction (National Reading Research Center). Reading Teacher, 49(3), 256-58.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. Journal of Educational Psychology, 88(1), 18-37.

Coladarci, T., & Gage, N. L. (1984). Effects of a minimal intervention on teacher behavior and student achievement. American Educational Research Journal, 21(3), 539-555.

Conley, M. M. W. (1983). Increasing students' reading achievement via teacher inservice education. Reading Teacher, 36(8), 804-808.

Copeland, W. D., & Decker, D. L. (1996). Video cases and the development of meaning making in preservice teachers. Teaching and Teacher Education, 12(5), 467-481.

Duffy, G. G., Roehler, L., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: a study of reading teacher effects. Reading Research Quarterly, 21(3), 237-252.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. Reading Research Quarterly, 22(3), 347-368.

Dupuis, M. M., Askov, E. N., & Lee, J.W. (1979). Changing attitudes toward content area reading: the content area reading project. Journal of Educational Research, 73(2), 65-74.

Greenberg, K. H., Woodside, M. R., & Brasil, L. (1994). Differences in the degree of mediated learning and classroom interaction structure for trained and untrained teachers. Journal of Classroom Interaction, 29(2), 1-9.

Hoover, N. L., & Carroll, R. G. (1987). Self-assessment of classroom instruction: an effective approach to inservice education. Teaching and Teacher Education, 3(3), 179-191.

Johnson, C. S., & Evans, A. D. (1992). Improving teacher questioning: A study of a training program. Literacy Research, 10.

Klesius, J. P., Searls, E. F., & Zielonka, P. (1990). A comparison of two methods of direct instruction of preservice teachers. Journal of Teacher Education, 41(4), 34-44.

Levin, B. B. (1995). Using the case method in teacher education: The role of discussion and experience in teachers' thinking about cases. Teaching and Teacher Education, 11(1), 63-79.

Miller, J. W., & Ellsworth, R. (1985). The evaluation of a two-year program to improve teacher effectiveness in reading instruction. Elementary School Journal, 85(4), 485-496.

Morrison, C., Harris, A. J., & Auerbach, I. T. (1969). Staff after-effects of participation in a reading research project: a follow-up study of the craft project. Reading Research Quarterly, 4, 366-395.

Olson, M. W., & Gillis, M. (1983). Teaching reading study skills and course content to preservice teachers. Reading World, 23(2), 124-133.

Reid, E. R. (1997). Exemplary Center for Reading Instruction (ECRI). Behavior and Social Issues, 7(1), 19-24.

Scheffler, A. J., Richmond, M., & Kazelskis, R. (1993). Examining shifts in teachers' theoretical orientation to reading. Reading Psychology, 14(1), 1-13.

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1996). Effects of introducing classroom performance assessments on student learning. Educational Measurement: Issues and Practice, 15(3), 7-18.

Spanjer, R. A., & Layne, B. H. (1983). Teacher attitudes toward language: Effects of training in a process approach to writing. Journal of Educational Research, 77(1), 60-62.

Stallings, J., & Krasavage, E. M. (1986). Program implementation and student achievement in a four-year Madeline Hunter follow-through project. Elementary School Journal, 87(2), 117-138.

Stallings, P., Robbins, P., Presbrey, L., & Scott, J. (1986). Effects of instruction based on the Madeline Hunter model on students' achievement: findings from a follow-through project. Elementary School Journal, 86(5), 571-587.

Streeter, B. B. (1986). The effects of training experienced teachers in enthusiasm on students' attitudes toward reading. Reading Psychology, 7(4), 249-259.

Talmage, H., Pascarella, E. T., & Ford, S. (1984). The influence of cooperation learning strategies on teacher practices, student perceptions of the learning environment, and academic achievement. American Educational Research Journal, 21(1), 163-179.

Tyre, B. B., & Knight, D. W. (1972). Teaching word recognition skills to preservice teachers: An analysis of three procedures. Southern Journal of Educational Research, 6(3), 113-122.

Wedman, J. M., Hughes, J. A., & Robinson, R. R. (1993). The effect of using systematic cooperative learning approach to help preservice teachers learn informal reading inventory procedures. Innovative Higher Education, 17(4), 231-241.

Wedman, J. M., & Moutray, C. (1991). The effect of training on the questions preservice teachers ask during literature discussions. Reading Research and Instruction, 30(2), 62-70.

Wedman, J. M., & Robinson, R. (1988). Effects of a decision-making model on preservice teachers' decision-making practices and materials use. Reading Improvement, 25(2), 110-116.

Westermark, T. I., & Crichlow, K. A. (1983). The effect of theoretical and situational knowledge of reading on teachers' estimates of readability. Reading Psychology, 4(2), 129-139.

Wham, M. A. (1993). The relationship between undergraduate course work and beliefs about reading instruction. Journal of Research and Development in Education, 27(1), 9-17.

Appendix B

Search Details

Search Terms Used and Number of Articles Returned: November 12, 1998

Key Term	OCLC - World Cat	OCLC - Article First	PsycINFO	ERIC
Reading teacher education		4	4	1,350
Teacher education reading	558	4	5	
Preservice reading teachers				
Training reading teachers	733	4	5	

Search Terms Used and Number of Articles Returned: April 13, 1999

SEARCH PsycLIT 1887- 1999	Teacher Education	Preservice Teacher Education	Inservice Teacher Education	Teacher Training	Teacher Preparation	Inservice	Preservice	Teacher Education Program Evaluation	Staff Development	Professional Development
Development										
Teacher Education	3,562									
Preservice Teacher Education	33	33								
Inservice Teacher Education	625	1	625							
Teacher Training	541	1	89	1,181						
Teacher Preparation	174	1	6	28	319					
Inservice	709	3	625	118	17	1,704				
Preservice	366	33	35	52	5	165	885			
Teacher Education Program Evaluation	2	0	0	0	0	0	0	2		
Staff Development	84	0	53	14	3	75	17	0	445	
Professional Development	138	5	39	12	8	106	35	0	90	1947
Reading	213	5	44	85	27	84	76	0	25	51
Writing	94	5	9	19	5	27	35	0	9	43
Literacy	55	3	6	11	9	25	26	0	6	22

Search Terms Used and Number of Articles Returned
Additional Searches:

Professional development <and> teacher - 247
Reading <and> inservice <and> teacher education - 52
Reading <and> writing <and> literacy -438
Reading <and> preservice <and> teacher education -35
Reading <and> writing <and> literacy <and> teacher education -14
Reading <and> writing <and> literacy <and> teacher education program evaluation - 0
Reading <and> writing <and> literacy <and> teacher training -1

Appendix C

Coding of Studies

Table 1: Preservice Studies

Author, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/In-service	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Copeland, W.D., & Decker, D.L. (1996). Teaching and Teacher Education,12(5), 467-481	Q	No	Yes	Pre 9	Group-based discussion experience with video of DRTA with 4th grade students 6 weeks	4*	Tr: How teachers interpreted (meaning-making) video data & improved after group discussion. Analyzed by topics. St: None	More than one-third of the topics discussed were adopted, transformed, or created by the respondents to describe their own meaning-making 3 weeks later.
Johnson, C.S., & Evans, A.D. (1992). Literacy Research, 10.	Q	No	Yes	Pre 22	Questioning techniques: a) shared inquiry; b) basal reader questions c) question & answer relationships (QAR). Six class sessions of 100 minutes = 100 minutes	Ele	Tr: Type of questions designed by trs after training. St: None	Number & percentage of restricted thinking and literal questions decreased while related and extended thinking questions (esp. critical thinking) increased.
Klesius, J.P. Searls, E.F., & Zielonka, P. (1990). Journal of Teacher Education,41(4), 34-44.	E	Yes + Random	No	Pre 37	Directed Reading Activity (DRA) [vocabulary, background, & motivation, guided silent reading, & comprehension questions] instruction via lecture & discussion vs instruction via video & simulation 1 semester (inferred)	Ele	Tr: Classroom observations of teachers St: None	There were no short-term differences in performance between the 2 instructional groups, but those instructed with videotape and simulation retained and used the information better for a longer period of time.
Olson, M.W., & Gillis, M. (1983). Reading World, 124-133.	E	Yes + Random assign--ment of teachers in existing program	No	Pre 121 n = ? attrition not reported	Learning study skills and content (topic: fundamentals of reading) concurrently. 1 semester (inferred)	Not stated	Tr: 1) Survey of Study Habits & Attitudes 2) Vocabulary: Stanford Diagnostic Reading Test 3) Compre: Stanford DRT 4) Researcher-designed MCQ course content test St: None	Experimental group had significant gains on survey of study habits and attitudes and course content. The control group tutored children in an elementary school. This suggests that preservice teachers lack proficient reading habits but can improve with lessons on study skills taught at the same time as content instruction.

Table 1: Preservice Studies (continued)

Tyre, B.B. & Knight, D.W. (1972). Southern Journal of Educational Research, 6(3), 113-122.	E	Yes. 3 Groups + random	Yes	Pre 72	Teaching word-recognition skills G1: classroom instruction supplemented with compiling an annotated list of teaching techniques G2: instruction supplemented with basal readers, preparing lesson plans, & viewing videos of teaching G3: studying basals, preparing lesson plans, & teaching children. Duration not reported.	Ele (inferred)	Tr: 1) Phonics Test for Teachers 2) Test of Phonics Principles 3) California Phonics Survey St: None	Phonics Test for Teachers and Test of Phonics Principles indicated significant gain in knowledge of word recognition skills for all groups. No significant gain reported when measured by California test. This supports the use of tutorial or practicum experiences together with learning of theory in preservice methods courses.
Wedman, J.M., Hughes, J.A., & Robinson, R.R. (1993). Innovative Higher Education, 17(4) 231-241.	E	Yes. Assigned to treatment & control based on GPA. Control n=30; treatment n=47	No	Pre 77	Using cooperative learning approach to help preservice teachers learn Informal Reading Inventory (IRI) concepts and procedures. 3 weeks.	1-6	Tr: 2 measures a) Researcher-designed MCQ based on Analytical Reading Inventory (Woods & Moe, 1981). Administered end of 3 weeks. b) Questions on participants' perceptions of the systematic cooperative learning approach. Treatment group only. St: None	Treatment group scored significantly higher than control group on IRI learning outcomes measure. The lecture method, used alone, appears significantly less effective for helping students learn skills and procedures. But cooperative learning alone is not sufficient as 32% of the participants perceived listening in class as an effective way to learn. A balance between traditional lecture method and various cooperative learning activities should be used.
Wedman, J.M., & Moutray, C. (1991). Reading Research and Instruction, 30(2), 62-70	E	Yes. 3 groups (A, B, C). Assignment by GPA. Equal numbers of low, middle, & high. Probably randomized	No	Pre 36	Questioning: Literal, text-based inferential questions (TBIQ) and Prior Knowledge-based inferential questions (PKBIQ) A: Literal & inferential questions taught. B: Lower & higher level questions in basals were taught. C: No instruction in questioning. Instruction on classroom uses of literature. 8-week project.	Jr ele	Tr: Written questions & audio record of questions generated during discussions St: None	Both Groups A & B generated TBIQ with higher overall frequency & percentage than C. As more practice was provided, frequency of TBIQ increased for all groups' questions. Group B performed significantly higher than A or C for TBIQ. All groups formulated PKBIQ. Group C asked more general questions than A or B.

Table 1: Preservice Studies (continued)

Wedman, J.M., & Robinson, R. (1988). Reading Improvement, 25(2), 110-116.	Q	No	Yes	Pre 27	Teachers' knowledge of reading and its effects on decisionmaking practices and awareness of materials. 1 semester (inferred)	Ele	Tr: Measures of decisionmaking (problem-solution and analysis by "thought units" and perception) (3-point scale). St: None	Participants' decision-making practices and materials use changed from pretest to posttest. Posttest thought units related to decisions addressed practices requiring instructional reevaluation or change. The number of materials mentioned substantially increased from pretest to posttest.
Wham, M.A. (1993). Journal of Research & Development in Education, 27(1), 9-17.	Q	No	Yes	Pre	Methods of teaching reading K-8. Undergraduate coursework + student teaching experience with cooperating teachers. N=35 Final phase of the study examined changes in orientation from pre-coursework to post-student teaching. Duration: about 3 semesters	K-8	Tr: DeFord's Theoretical Orientations to Reading Profile (TORP) 6/35 teachers were videotaped to ensure consistency between reading instruction and responses to TORP. St: None	More than half (54%) did not experience any change in theoretical orientation throughout the course. This finding appears to support the view (Lortie, 1975) that the major influence shaping future teachers' conceptions of teaching is past experience as pupils. The current study suggests that the methods courses and the student teaching experience are only modestly related to changes in the theoretical orientations of preservice teachers.

Table 2: Studies of Both Preservice and Inservice

Authors, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type Of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Levin, B.B. (1995). Teaching and Teacher Education,11(1) - 63-79.	Q Participants from existing program.	Yes. Not reported whether there was random assignment. Each group had equal numbers of student, beginning, and experienced teachers.	Yes	8 pre 16 in	Case method in teacher education. 2 cases: teaching writing in 4th grade Reading & writing about case vs. reading, discussing, & writing. About 5 weeks	4*	Tr: analyses of written response to cases St: None	Opportunity to read, write, and discuss a case affected teachers' thinking about the case. For experienced teachers, discussion was a catalyst for reflection and metacognition. Less experienced teachers and preservice teachers were able to clarify and elaborate their thinking. Only reading and writing about a case provided little stimulus for teachers to elaborate their understanding or increase their perspectives on issues in the case.
Westermark, T.I., & Crichlow, K.A. (1983). Reading Psychology: An International Quarterly, 4, 129-139.	Q	Yes. Not random: 4 groups (1, 2, 3, and 4) from existing enrollment	No	72: 36 pre + 36 in	Estimating readability levels G1: Theoretical & situational knowledge of reading G2: Situational knowledge only G3: Theoretical knowledge only G4: No knowledge of theory or practice. 1 semester (inferred)	Ele	Tr: Accuracy in estimating readability subjectively compared to actual reading level of materials. St: None	No effect. Teachers vary widely in estimating readability. All groups estimated readability equally accurately, and accuracy decreased as readability levels increased for all groups.

Table 3: Inservice Studies with Teacher Outcome Measures Only

Author/s, Date & Pub	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/-No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Dupuis, M.M., Askov, E.N., & Lee, J.W. (1979). Journal of Educational Research, 73(2), 65-74.	Q	Yes. Non-random assignment. Control group consisted of teachers from the same school but not part of the project.	Yes; teachers only.	In 127	Teaching content area teachers how to teach reading (diagnosis, motivation, organization for instruction, materials selection, skills development, evaluation, etc.). Also aimed to improve attitudes toward teaching reading in the content area classroom. Videotapes used. Duration: 1 year. Training hours: about 45 hours.	Jr high	Tr: Teacher attitude toward teaching reading, teacher morale, teacher skill level in teaching reading, and staff ratings of teacher change. St: None.	The attitude gains made by the experimental groups were significantly greater than those of the comparison groups. Morale appeared not to have been a significant factor in determining teachers' attitudes to integrating reading instruction in content areas. Significantly more experimental group teachers changed from nonmastery to mastery at posttest. The interrater reliability of judging teacher change cannot be determined. Ratings (observations) seemed to indicate that changes were reflected in classroom practice. But overall, teachers' knowledge of reading skills did not improve as much as hoped.
Greenberg, K.H., Woodside, M.R., & Brasil, L. (1994). Journal of Classroom Interaction, 29(2), 1-9.	Q	Yes. Not random. Existing classrooms used.	No	In 27	Questioning & teacher-student interactions (tr- questions, st-answers, tr-sustaining/ terminating feedback). COGNET: Cognitive Enrichment Network. Explored relationship between mediated learning interactions (based on Vygotsky & Feuerstein) and question dyad variables. Duration: 3 years. More than 60 hours of training.	K, 1, 2, 3	Tr: Observational analysis based on Mediated Learning Experience Observational Analysis System (Greenberg, 1990); Brophy & Good's Teacher-Child Dyadic Interaction System (1969). St: None	The degree to which teachers provided mediated learning was based on knowledge and skills on how to facilitate the learning process. Experimental group showed higher levels of use of mediated learning, e.g., asking more process questions and accepting partially correct answers. They were able to ask questions requiring children to choose between responses and encouraged them to think more deeply through rephrasing and giving clues. Limitation: small sample.

Table 3: Inservice Studies with Teacher Outcome Measures Only (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Hoover, N.L., & Carroll, R.G. (1987). Teaching & Teacher Education, 3(3), 179-191.	Q	No	Yes; teachers only	In 53	Training in using reading groups was provided. Teachers were also trained to use a self-assessment checklist to evaluate their own reading instruction on videotape. Duration: about 6 months. 32 hours of training.	K-7	Tr: Random selection of video taped data (37%) was made and audited by the researchers using the self-assessment checklist. St: None	Pre- and posttest data showed that the self-assessment procedure was effective in helping teachers improve instruction. Teachers reported significant improvements in their teaching behavior, which is supported by the quantitative data. Unanswered question: whether the impact of self-assessment procedure is sustainable.
Morrison, C., Harris, A.J., & Auerbach, I.T. (1969). Reading Research Quarterly, 4, 366-395.	Q	Yes	Yes	In 92	Project was designed for teaching beginning reading instruction to disadvantaged urban children. Two approaches consisting each of two methods: a) skills approach (basal reader & phonovisual method); b) Language Experience Approach (LEA & LE Audio-Visual; d) Pilot (combination of LEA & word recognition). Duration: 3 years. Training hours: not reported.	1, 2, 3	Tr: Attitude inventory & interviews St: N = 1378 (started); replication study, N = 679 St: None	Results indicated that teachers were no longer using the experimental materials in the same way they did when the study was in progress and had, in fact, reverted to a pattern of instruction similar to what they had been doing. Administrative policies pertaining to grouping also reverted back to what they were. The study cautions against using short-term results to validate teacher education efforts.
Scheffler, A.J., Richmond, M., & Kazelskis, R. (1993). Reading Psychology: An International Quarterly, 14(1), 1-13.	Q	No	Yes. Teachers only. Pre-test, post-test, and delayed--posttest	In 55	Whole language. Duration: 2.5 months 2 day-long workshops. About 12-16 hours of training.	K-8	Tr: Theoretical orientation to reading as measured by the DeFord Theoretical Orientation to Reading Profile (TORP) St: None	A significant main effect was found among the pre-, post-, and delayed-post-trial scores for the total TORP scores. As a group, the subjects moved significantly closer to a whole language orientation from the pre- to the delayed-post-trial measure.

Table 3: Inservice Studies with Teacher Outcome Measures Only (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Spanjer, R.A., & Layne, B.H. (1983). Journal of Educational Research, 77(1), 60-62.	Q	No	Yes. Teachers only	In 78	Teaching writing as a process to facilitate a change in teachers' attitude to language. Workshop curriculum adapted from Berkeley's Bay Area Project (1977). Duration: 3 years. 3 workshops over 3 summers. Training hours not reported.	38 elem.-mid-sch.; 41 sec.-post-sec.	Tr: Teachers' attitudes were assessed using the Language Inquiry (Frogner, 1969) inventory. The instrument covers standards in using American English. and on language study & teaching. 1 missing pretest score, n=78 St: None.	The posttest mean was significantly greater than the pretest mean. The process approach to writing may influence teachers' attitudes toward language (i.e., less rule bound & prescriptive, more sensitive to usage according to purpose and context).

Table 4: Inservice Studies with Teacherand Student Outcome Measures

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Anderson, L.M., Evertson, C.M., & Brophy, J.E., (1979). Elementary School Journal, 79(4), 193-223.	Q	Yes. Not truly random. 10 treatment (observed); 10 control; 7 treatment (not observed). All in each school assigned to control or treatment.	No	In 27	Instructional model for promoting effective instruction in small groups in the early grades. Duration: 1 year. Minimal training. Teachers read a manual and 2 meetings were held. Training hours: None.	1	Tr: Observations over 1 year to ensure implementation of the model. St: Class means were reported. 27 classes. Measures of student readiness (Metropolitan Readiness Tests, Level 1) and reading achievement.	The treatment classes (whether observed or unobserved) had higher adjusted achievement scores. Differences in teachers' behaviors in the control and experimental groups were observed, but not all can be attributed to the treatment. The treatment teachers exhibited more of those behaviors associated with achievement. Overall, the content of the treatment probably had effects on student achievement, but other effects (e.g., school effects) cannot be ruled out.
Baker, J.E., (1977). Ontario Psychologist, 9(4), 57-62.	Q	Yes (students only). Not random. N = 18 (control) and N = 18 (treatment)	Yes; students only	In 18	Classroom consultation model (IS/C) to improve reading instruction for underachieving readers. Strategies include: 1) stimulus variation 2) reinforcement techniques 3) response guidance cues 4) questioning techniques. Videotapes (of elementary & secondary teachers & their students) used for training. Duration: 4.5 months. 10 workshops (+ 6 attended previously) About 10-15 hours of training	4	Tr: N = 18. Teachers' ratings of relevance of the inservice sessions and written evaluations, indicating changes in attitudes, values and behavior. St: N = 36 (underachieving readers taught by 3 teachers) Dependent measures: Gilmore Oral Reading Test (Accuracy & Comprehension subtests); Schonell Graded Word Reading List; Metropolitan Achievement Test (MAT): Elementary Spelling Subtest.	Changes in the teachers included increased awareness of questioning techniques, improvement in planning skills and introducing concepts sequentially, requiring and illustrating principles of thinking, utilizing oral discussions to encourage student writing, allowing time for concrete presentations of concepts. The results were significant in student performance for three of the four dependent measures. The MAT (spelling test) did not show significant differences between treatment and control subjects.

Table 4: Inservice Studies With Teacher and Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Book, C.L., Duffy, G.G., Roehler, L.R., Meloth, M.S., & Vavrus, L.G. (1985). Communication Education, 34(1), 29-36.	E	Yes. Randomly assigned.	No	In 22	Direct Explanation method. Teachers were trained in the use of explicit explanations of reading skills and processes. Low reading groups. Duration: not reported. 3 training sessions. Number of training hours not reported.	5	Tr: Teachers in control and experimental groups observed and rated on explicitness of their explanations, using a rating scale developed by the researchers. St: After each lesson, at least 5 students were interviewed on strategy awareness: - What did you learn? - Why is it important? - How do you do it? No measures of student strategy usage or reading achievement.	Students of treatment teachers scored significantly higher than students of control teachers on strategy awareness. Treatment teachers were rated as significantly more explicit in their explanations than control teachers. Treatment teachers also became more explicit in their explanations over time. There was a significant positive relationship between student metacognitive awareness and teacher explanation, i.e., as teachers became more explicit in their explanation, students' ratings of awareness increased.
Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). Journal of Educational Psychology, 88 (1), 18-37	Q	Yes. Teachers were not randomly assigned.	Yes (for students)	In 10 teachers 60 students	Transactional Strategies Instruction (TSI), emphasizing joint construction of text interpretations and student strategy usage. Students read below 2nd grade level. Duration: 1 academic year. Training hours: not reported. The experimental group teachers did not receive training for this study. All had extensive prior experience with TSI.	2	Tr: No formal measures were used although lessons were observed. Treatment classes were observed to have more prominent discussion of strategies than comparison reading groups. St: a) Strategies interview to assess awareness of comprehension and problem-solving strategies. b) Retelling questions to assess students' retelling and sequencing of 2 stories. c) Think-aloud task to determine whether students were more text- or reader-based in their responses to probes. d) Standardized subtests of reading comprehension and word skills (Stanford Achievement Test [SAT]).	Students of treatment teachers scored significantly higher than students of control teachers on the comprehension and word skills subtests of SAT. They also showed significantly greater improvement on these measures over the course of the study. Students of the treatment teachers recalled more literal information and were significantly more interpretive in their retelling of the stories. Students of treatment teachers reported more awareness of comprehension and word-level strategies than did the students of control teachers.

Table 4: Inservice Studies With Teacherand Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Coladarci, T., & Gage, N.L. (1984). American Educational Research Journal, 21(3), 539-555.	E	Yes	Yes; pre and post for teachers & students. 28 classes (data available)	In 32	Teacher education packets (TEP) by Crawford et al., 1978 were given to treatment and control group teachers. TEP consisted of a) behavior management & discipline b) large-group instruction, use of Q&A, & phonics exercises in reading; c) questioning and feedback strategies. There was no formal training; teachers were asked to follow what was given in the guidelines. Duration: About 1 year. Formal training: None.	4,5,6	Tr: Classroom observation pre- and posttreatment. Observation records yielded rough estimates of the extent to which teacher behavior reflected TEP recommendations St: Comprehensive. Test of Basic Skills was used.	As an experiment, this study failed to corroborate the positive results obtained previously in similar classroom-based experiments. Toward the end of the school year, the experimental group teachers did not show appreciably greater conformity to the TEP recommendations, nor did their classes improve in end-of-year academic achievement.
Conley, M.M.W. (1983). Reading Teacher, 36(8), 804-808.	E	Yes. Teachers in each school were randomly assigned	Yes, for students only	In 32	Comprehension instruction (literal, inferential, critical, and creative). Included higher order questioning techniques. Note: Students were all black & from low socioeconomic backgrounds. They were selected because they read below the national norm for their age level. Duration: 6 months. About 10-15 hours of training.	Grade 6 materials were used; students were ungraded	Tr: Ongoing formative evaluation. St: Gates-MacGinitie Reading Test (level E)	Teachers benefited (evident from qualitative data during inservice evaluations and feedback), but more important were significant comprehension gains for students.

Table 4: Inservice Studies With Teacher and Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Duffy, G.G., Roehler, L.R., Meloth, M.S., Vavrus, L.G., Book, C., Putnam, J., & Wesselman, R. (1986). Reading Research Quarterly, 21(3), 237-252.	E	Yes, randomly assigned	1) Yes: Baseline data on teacher effectiveness through observations + post-treatment observations. 2) Yes: students were measured pre- and post- on standardized test.	In 22	Explicit instruction and explanation in using reading skills strategically. Low reading groups. Duration: 7 months. 1 meeting & presentation + 10 hours of training.	5	Tr: Ratings of teachers' instructional explanations (transcripts) St: Ratings of "awareness" after lessons (transcripts): 5 students interviewed per teacher. Gates-MacGinitie Reading Test (1978) Time taken for control and treatment groups to do the test.	Teachers who were trained were rated significantly higher than those in the control group in explicit strategy instruction. Students in the experimental groups showed significantly more awareness of reading strategies. But there were no achievement gains in comprehension. The qualitative data from 3 good teachers and 3 less effective teachers showed the former producing significantly greater growth in achievement. Students in the treatment group took longer to complete the posttest.
Duffy, G.G., Roehler, L.R., Sivan, E., Rackliffe, G., Book, C., Meloth, M.S., Vavrus, L.G., Wesselman, R., Putnam, J., & Bassin, D. (1987). Reading Research Quarterly, 22 (3), 347-368.	E	Yes. Randomly assigned	Yes	In 20 148 students	Direct Explanation with a focus on explaining the reasoning associated with skill and strategy usage. Low reading groups. Duration: 1 academic year. Training hours: 12	3	Tr: Researcher-designed rating instrument was used to rate transcripts of teachers' explanations for explicitness. St: a) SAT (comprehension & word skills subtests) b) Michigan Educational Assessment Program (MEAP) [delayed posttest] c) Lesson interviews (immediately following a lesson) & concept interviews (at the end of the year) d) Supplemental Achievement Measure (SAM) [researcher-designed] e) Modified Graded Oral Reading Paragraph (GORP)	The treatment teachers were found to be more explicit in explaining the reasoning associated with using reading skills than were the control teachers. On SAT, students of treatment teachers scored significantly higher than students of control teachers on word skills, but not on comprehension. Students of treatment teachers scored significantly higher than students of control teachers on MEAP. Students of treatment teachers scored higher also in -Lesson interviews -Concept interviews -SAM (Part 2 only, not Part 1) -Modified GORP test.

Table 4: Inservice Studies With Teacher and Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Miller, J.W., & Ellsworth, R. (1985). Elementary School Journal, 85(4), 485-496.	Q & E	Yes a) Not random for N = 143 b) Random for N = 33	1) Yes for teachers 2) No complete pretest data for students, but baseline was established through pretest scores of equivalent students.	In 141/143	Four semester-long courses aimed at improving reading instruction: a) assessment of reading levels & skill need b) differentiation of instruction c) use of diverse instructional materials d) Directed Reading Activity (DRA) as basic format for lesson preparation; e) story discussion techniques f) promotion of recreational reading & developing student reading interests. Duration: 2 years. Training hours: not reported.	2-5	Tr: a) Knowledge of reading assessed by the Inventory of Teacher Knowledge of Reading (Artley & Hardin, 1975). N = 143. Not random. b) Measurement of actual teacher behavior. Classroom observation. N = 16 (exp). N = 17 (control). Random. St: California Achievement Test. (N = 511). Post-inservice training program comparison of participating and non-participating teachers' students.	Teachers who had more knowledge of reading, but less experience and fewer college degrees, opted to participate in the inservice course. Teacher attitudes toward reading instruction showed significant differences on three (adjusted) posttest means: 1) grouping children on the basis of interests has no place in a reading program (trained teachers disagreed more strongly); 2) if a child does not respond to phonics instruction, he should be taught to read by sight (trained teachers agreed more strongly); 3) reading is a skill and must be practiced if proficiency is to be achieved (trained teachers agreed more strongly). Trained teachers (N = 16) demonstrated higher implementation levels of desired behaviors in all six areas than did a sample of nonparticipating teachers (N = 17). A posthoc analysis showed that there were significant differences in student achievement at 0.05 level.
Stallings, J., Robbins, P., Presbrey, L., & Scott, J. (1986). Elementary School Journal, 86(5), 571-587.	Q	Yes. Not random.	1) Yes, teachers were observed before training. 2) Yes for students	In 13 teachers; 208 students	Madeline Hunter's Instructional Theory into Practice to improve instruction and classroom management. Funding was given by NIE to improve reading and math of Chapter 1-eligible children. 2 selected schools had the highest percentage of Chapter 1-eligible children in their school districts (50% & 55%). Reports data from 1982-1983 (II), 1983-1984 (III). Duration: 3 years. Training hours: not reported.	1-4	Tr: Quality and quantity of program implementation were measured by Instructional Skills Observation Instrument (ISOI) & Time-Off-Task Observation System, questionnaires & interviews. Designed by the researchers. St: a) Reading and math achievement scores. b) Rate of student engagement as measured by Time-Off-Task system.	Teachers improved in their instructional skills significantly over 4 months. The range in teacher performance was reduced. Students made significant gains in reading during Phases II & III of the study and in math during Phase II, but instructional skills and engaged rate did not correlate with gain. Limited English-speaking (LES) students benefited from the program. Their gains each year in reading and math were more than those of the other children in the study.

Table 4: Inservice Studies With Teacher and Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Stallings, J., & Krasavage, EM (1986). Elementary School Journal, 87(2), 117-138.	Q	Yes. Not random	Yes for teachers & students	450 students	As above. Reports data from 1984-5 (IV). Schools selected had the highest percentage of Chapter 1-eligible students. Duration: as above. Training hours: as above.	1-4	As above	Seven of ten teachers' reading and math ISOI scores dropped in 1985. Student engaged rates in reading and math dropped significantly in 1985. Comparisons with matched control schools on standardized tests showed greater gains among control students from 1984-1985. LES students gained more than English-speaking students. Inconsistencies in teacher behaviors and student reading achievements in all years of the study. Evidence is not strong for a link between implementation of the Hunter model and student achievement.
Streeter, B.B. (1986). Reading Psychology: An International Quarterly, 7(4), 249-259	E	Yes. Teachers were randomly assigned	Yes	In 19	Enthusiasm training for teachers. Videotapes used for postconferencing. Duration: 2 weeks, 10 hours of training.	1-5	Tr: Teachers were observed pre- and posttraining. Variables include vocal delivery, eyes, gestures, movements, facial expressions, word selection, acceptance of ideas, and overall energy. St: Attitudes to reading measured by the SRA Primary Level (pre and post).	The control group showed some gains, but not as much as the experimental group. Training led to increased levels of observable teacher enthusiasm. Only one of the four dimensions of the student measure showed significant change. There was a drop in the "Expressed Reading Difficulty" dimension, showing less perceived difficulty with reading. Hence, teachers' higher levels of enthusiasm posttraining had an effect on students' attitudes to reading.

Table 4: Inservice Studies with Teacher and Student Outcome Measures (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control/ Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Talmage, H., Pascarella, E.T., & Ford, S. (1984). American Educational Research Journal, 21(1), 163-179.	Q	Yes. Not random	Yes. Teachers and students (except 1st grade)	In 107	Increasing teachers' skills in conducting cooperative learning activities. Duration: 3 years. Training hours: not reported.	2-6	Tr: Classroom observations, interviews and pre- and postmeasures of teaching practices and teacher attitudes were used. St: 1) Students' perceptions of their classroom learning environment were obtained. 2) Reading Comprehension & Language Arts achievement measured by district standardized tests (Science Research Associates, Inc.)	Cooperative learning strategies can be learned effectively by teachers through long-term inservice programs. There were positive effects of teacher experience with cooperative grouping on student perceptions of cooperation. There were some effects on students' reading scores but not for language arts. Some ambiguity still exists in accounting for the influence of cooperative learning on achievement. There are probably other unmeasured outcomes of the project that helped raise student achievement.

Table 5: Inservice Studies With Student Outcome Measures Only

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Block, C. (1993). Elementary School Journal, 94(2), 139-151.	E	Yes. Classes were randomly assigned.	Yes (for students)	Research assistants were used. No. not reported. 352 students	Strategy instruction in a student-centered curriculum, i.e., student choice of objectives and materials. Duration: 8 months. Training hours: not reported.	2-6	Tr: None St: a) Iowa Test of Basic Skills was administered posttest. It was not reported whether this was used at pretest. b) Observations: The last lesson taught in each experimental and control class was videotaped and rated for levels of comprehension and thinking abilities seen in discussions. c) Students' self-esteem, idea generation, and reflective thinking ability were assessed pre- and posttest. d) Reasoning ability was measured using the California State Department of Education Statewide Assessment Test (1989).	Experimental students scored significantly higher than controls on the posttests for reading comprehension, vocabulary, and total battery scores. No significant differences were found between the two groups' scores on the English grammar posttest. On the basis of videotaped lesson observations, raters ranked students in experimental classes as "better thinkers" than controls. Experimental students did better than controls on measures of self-esteem, idea generation, ability to transfer thinking skills to real-life situations, reflective thinking, reasoning, and problemsolving.
Brown, R., El-Dinary, P.B., Pressley, M., & Coy-Ogan, L. (1995). Reading Teacher, 49(3), 256-258	Q	Yes. Not random: used teachers from existing classrooms	No	In 10 teachers 12 students	Students "were experiencing at least some difficulty learning how to read." 1-year study. Training hours: not reported.	2	Tr: None St: Several measures of reading & strategic processing (instruments not stated) (ref: Brown et al. Tech report). Instruments are described in Brown, Pressley, Van Meter, & Schuder (1996), Journal of Educational Psychology, 88(1), 18-37.	TSI students: a) learned more about strategic processing and used strategies on their own more frequently while reading a challenging story; b) acquired more information and developed richer understanding from stories read; c) showed greater gains on standardized comprehension and word study skills. Teachers believed it increased students' self-confidence and enjoyment as readers, improving interactions among students during reading. Teachers also found it challenging to teach students to use a repertoire of strategies.

Table 5: Inservice Studies With Student Outcome Measures Only (continued)

Author/s, Date, & Publisher	Exp/ Quasi	Control: Yes/No	Pre & Post: Yes/No	Pre/ In-S	Type of Teacher Training & Duration	Gr	Dependent Measures: Teacher (Tr) & Student (St)	Findings
Reid, E.R. (1997). Behavior and Social Issues, 7(1), 19-24.	Q	Yes. Not random	No	In N not stated	Training for a language arts/integrated curriculum: word recognition, vocabulary comprehension, study skills, spelling, penmanship, proofing, writing, and literature. Training included the above, using strategies that prevent failure and management systems to enable all students to learn. Microcomputers used to teach typing, reading, and spelling in K-8. Duration: 1 year. 5-day seminar. Approximately 30-35 hours.	1-12	Tr: None St: SAT, CTBS, & ITBS. Woodcock-Johnson & Nelson-Denny (for some of the special ed and bilingual students in two schools). Included regular education, special ed, gifted, and special needs students. 2,274 students (1990); regular students N = 1,733. 1,986 students (1996).	In the 1990 evaluation, looking only at the schools with controls, the experimental schools gained 8 & 14 Normal Curve Equivalents (NCEs) in vocabulary and comprehension compared with a range from a loss of 9 NCEs to a gain of 6 NCEs for control schools. For the 1996 evaluation, students demonstrated significant gains on the reading subtests of standardized achievement tests.
Shepard, LA., Flexer, R.J., Hiebert, E.H., Marion, S.F., Mayfield, V., & Weston, T.J. (1996). Educational Measurement: Issues and Practice, 15(3), 7-18.	Q	Yes. Not random. Treatment schools volunteered and control schools were matched on SES data.	Approximately premeasures appropriate for 3rd graders used and compared with outcome scores at the end of the year.	In	Performance assessment in reading and math. After school workshops were held weekly for a whole year alternating between reading and math. Duration: 1 year. Training hours: not reported.	3	Tr: None St: 1991 Maryland School Performance Assessment Program, supplemented by a portion of another measure (Korets et al., 1991) for math. N = 335.	No gains in student learning were found following the yearlong effort to introduce classroom performance assessments.

Appendix D

Standards

The 1989 NCATE Approved Curriculum Guidelines of the ACEI for the basic programs for the preparation of elementary education teachers include the following standards: (Note that indicators are provided for Standard 13, the standard dealing with literacy.)

1. Programs should provide teacher candidates with an understanding of the roles of elementary school teachers and the alternative patterns of elementary school organization.
2. Programs should provide study and experience concerning the role of the teaching profession in the dynamics of curriculum change and school improvement.
3. Programs should include study and experiences, throughout the professional studies sequence that link child development to elementary school curriculum and instruction.
4. Programs should develop the teacher candidates' capacities to organize and implement instruction for students.
5. Programs should include study and application of a variety of developmentally appropriate experiences that demonstrate varied approaches to knowledge construction and application in all disciplines.
6. Programs should include study and application of current research findings about teaching and learning.
7. Programs should provide a well-planned sequence of varied clinical/field experiences with students of different ages, cultural and linguistic backgrounds, and exceptionalities. These experiences should connect course content with elementary school practice.
8. Programs include opportunities to study, analyze, and practice effective models of classroom management in campus and field-based settings and to engage in a gradual increase in responsibility.
9. Programs should provide study and experiences for critically selecting and using materials, resources, and technology appropriate to the age, development level, cultural and linguistic backgrounds, and exceptionalities of students.
10. Programs should provide for indepth study in at least one academic discipline by including significant course work beyond the introductory level to reflect processes of inquiry and research.
11. Programs should develop understandings of positive health behaviors, movement skills, and physical fitness to allow teacher candidates to provide appropriate health education and physical education experiences for students.
12. Programs should prepare teacher candidates to become confident in their ability to do mathematics and to create an environment in which students become confident learners and doers of mathematics.
13. Programs in the area of students' literacy development should be designed to help teacher candidates create experiences for their students in reading, writing, and oral language. These programs should stress the integration of reading, writing, and oral language with each other and with the content areas of the elementary school curriculum.

    Program emphasis include study of and experiences with:

13.1 The cognitive and linguistic foundations of literacy development in students

13.2 Ways of promoting vocabulary growth in students

13.3 The flexible use of a variety of strategies for recognizing words in print

13.4 Teaching of the conventions of language needed to compose and comprehend oral and written texts (e.g., text structure, punctuation, spelling)

13.5 The strategies readers can use to discover meaning from print and to monitor their own comprehension

13.6 The ways listening, speaking, reading, and writing relate to each other and to the rest of the elementary curriculum

13.7 Identifying and developing appropriate responses to differences among language learners (e.g., linguistic, sociocultural, intellectual, physical)

13.8 Communicating with parents concerning the school language program and developmentally appropriate language experiences at home

13.9 Speaking and writing that vary in form, subject, purpose, audience, point of view, tone, and style

13.10 Ways to promote reading, writing, and oral language for personal growth, lifelong learning, enjoyment, and insight into human experience

13.11 The literature of childhood including (a) knowing a range of books, (b) knowing how to share literature with students, and (c) knowing how to guide students to respond to books in a variety of ways

13.12 Promoting creative thinking and expression, as through storytelling, drama, choral/oral reading, imaginative writing, and the like.

14. Programs in science for teacher candidates should focus on academic, personal, social, and career applications of the biological, earth, and physical sciences and should develop skills in instruction to promote these understandings and positive attitudes among students and youth.

15. Programs should prepare teacher candidates to translate knowledge and data-gathering processes from history and the social sciences into appropriate and meaningful social studies experiences for students.

16. Programs should prepare teacher candidates to translate knowledge of and experience in the visual and performing arts into appropriate experiences for students.

The 1983 NCATE Approved Curriculum Guidelines of the International Reading Association for advanced programs in reading education follow in this report, but readers should be aware that IRA has published a 1998 revision of the standards for reading professionals. The 1998 standards will be applied to programs of institutions currently seeking accreditation or continuing accreditation.

Competencies required of candidates from those institutions presently approved are the following:

1. Philosophy of Reading Instruction: Reading is a complex, interactive, and constructive process.

1.1 Recognizes the importance of teaching reading as a process rather than as a discrete series of skills to be taught through unrelated activities/exercises

1.2 Recognizes the importance of using a wide variety of print throughout the curriculum, including high-quality children's/adolescents' literature and diverse expository materials appropriate to the age and developmental level of learners

1.3 Has knowledge of current and historical perspectives about the nature and purposes of reading and about widely used approaches to reading instruction

1.4 Recognizes and appreciates the role and value of language in the reading and learning processes

1.5 Recognizes the importance of embedding reading instruction in a meaningful context for the purpose of accomplishing specific authentic tasks or for pleasure

1.6 Recognizes the value of reading aloud to learners.

2. Professionalism

2.1 Pursues knowledge of reading and learning processes by reading professional journals and publications and participating in conferences and other professional activities

2.2 Employs inquiry and makes thoughtful decisions during teaching and assessment

2.3 Interacts and participates in decisionmaking with teachers, teacher educators, theoreticians, and researchers and plays an active role in schools, classrooms, and the wider professional community

2.4 Supports and participates in efforts to improve the reading profession by being involved in licensing and certification

2.5 Participates in local, state, national, and international professional organizations whose mission is the improvement of literacy

2.6 Promotes collegiality with other literacy professionals through regular conversations, discussions, and consultations about learners, literacy theory, and instruction

2.7 Shares knowledge, collaborates, and teaches with colleagues, as in inclusion programs.

3. Moral Dimensions and Values

3.1 Recognizes the importance of literacy as a mechanism for personal and social growth

3.2 Recognizes that literacy can be a means for transmitting moral and cultural values within a community

3.3 Recognizes values and is sensitive to human diversity

3.4 Recognizes and is sensitive to the needs and rights of individual learners.

4. Perspectives About Readers and Reading

4.1 Understands and accepts the importance of reading as a means to learn, to access information, and to enhance the quality of life

4.2 Understands and is sensitive to differences among learners and how these differences influence reading

4.3 Understands and respects cultural, linguistic, and ethnic diversity and recognizes the positive contributions of diversity

4.4 Believes that all students can learn to read and share in the communication process

4.5 Recognizes the importance of using reading in positive ways in the classroom

4.6 Recognizes the value and importance of creating a supportive and positive environment for literacy learning

4.7 Recognizes the importance of giving learners opportunities in all aspects of literacy as readers, authors, and thinkers

4.8 Recognizes the importance of implementing literacy programs designed to meet the needs of readers rather than imposing prescribed, inflexible programs

4.9 Recognizes the importance of building on the strengths of individual learners rather than emphasizing weaknesses.

5. Language, Development, Cognition, and Learning

5.1 Understands that language is a symbolic system

5.2 Understands major theories of language development, cognition, and learning and uses them to implement a well-planned and comprehensive reading program

5.3 Is aware of the linguistic, sociological, cultural, cognitive, and psychological bases of the reading process

5.4 Is aware of the physical, emotional, social, cultural, environmental, and intellectual factors on learning, language development, and reading

5.5 Understands dialect variations and respects linguistic differences

5.6 Understands the importance of language development in relation to reading and writing.

6. Knowledge of the Reading Process

6.1 Perceives reading as the process of constructing meaning through the interaction of the reader's existing knowledge, the information suggested by the written language, and the context of the reading situation

6.2 Is aware of relationships among reading, writing, listening, and speaking

6.3 Has knowledge of emergent literacy and the kinds of experiences that support literacy

6.4 Is aware that reading develops best through activities that embrace concepts about the purpose and function of reading and writing and the conventions of print

6.5 Understands the role of models of thought that operate in the reading process

6.6 Is able to explain the model various word recognition, vocabulary, and comprehension strategies used by fluent readers

6.7 Understands the role of metacognition in reading

6.8 Has knowledge of the importance for reading in language development; listening ability; cognitive, social, and emotional development; and perceptual motor abilities

6.9 Understands the nature and multiple causes of reading disabilities

6.10 Understands the relationship of phonemic, morphemic, semantic, and syntactic systems of language to the reading process.

7. Creating a Literate Environment

7.1 Promotes the development of a literate environment that fosters interest and growth in all aspects of literacy

7.2 Uses texts to stimulate interest, promote reading growth, foster appreciation for the written word, and increase the motivation of learners to read widely and independently for information and for pleasure

7.3 Models and discusses reading as a valuable activity

7.4 Engages students in activities that develop their image of themselves as literate

7.5 Promotes feelings of pride and ownership for the process and content of learning

7.6 Provides regular opportunities for learners to select from a wide variety of books or other quality written materials

7.7 Provides opportunities for students to be exposed to a variety of high-quality, relevant reading materials

7.8 Provides opportunities for students to be exposed to various purposes for reading/writing, to experience reading/writing as relevant to themselves, and to write and have their writing responded to in a positive way

7.9 Recognizes the importance of providing time for reading of extended text for authentic purposes

7.10 Provides opportunities for creative response to text.

8. Organizing and Planning for Effective Instruction—Knowledge of Contextual Factors

8.1 Understands how factors such as content, purpose, tasks, and settings influence the reading process

8.2 Provides flexible grouping based on students' instructional levels, rates of progress, interests, or instructional goals

8.3 Understands how assessment and grouping procedures can influence motivation and learning

8.4 Understands how environmental factors can influence students' performance on measures of reading achievement

8.5 Understands the relationship among home factors, social factors, and reading habits in students

8.6 Understands the influence of school programs (e.g., remedial, gifted, tracking) on students' learning

8.7 Understands the conditions necessary for all students to succeed.

9. Knowledge of Individual Differences

9.1 Understands what the reader brings to the reading experience (e.g., prior knowledge, metacognitive abilities, aptitudes, motivation, attitude)

9.2 Understands the influence of cultural, ethnic, and linguistic backgrounds on the reading process

9.3 Understands the relationship among reader's self-concept, attitudes, and learning

9.4 Understands the interactive nature and multiple causes of reading difficulties.

10. Knowledge of Instructional Materials

10.1 Understands how to design, select, modify, and evaluate materials that reflect curriculum goals, current knowledge, and the interests, motivation, and needs of individual learners

10.2 Understands the structure and content of various texts used for instruction

10.3 Understands and uses new instructional technologies

10.4 Understands methods for determining whether materials are clear and appropriate for individual students.

11. Knowledge of Instructional Strategies---Teaching Strategies

11.1 Provides direct instruction and models what, when, and how to use reading strategies with narrative and expository texts

11.2 Models questioning strategies

11.3 Employees strategies to encourage and motivate students to pursue and respond to reading and writing for personal growth and fulfillment

11.4 Teaches effective study strategies

12. Learning Strategies

12.1 Helps students learn and apply comprehension strategies for a variety of purposes

12.2 Helps students monitor their comprehension and reading processes

12.3 Understands and helps students learn and apply reading comprehension strategies in the content areas

12.4 Helps students gain understanding of the conventions of language and literacy

12.5 Teaches word recognition through the use of context, word analysis, and syntactic cueing strategies

12.6 Helps students learn that word recognition strategies aid comprehension

12.7 Helps students learn effective techniques and strategies for the ongoing development of vocabulary

12.8 Helps students analyze information presented in a variety of texts

12.9 Helps students connect prior knowledge with new information

12.10 Assists students in assuming control of their reading

12.11 Helps students use new technology and media effectively.

13. Demonstrate Knowledge of Assessment Principles and Techniques

13.1 Recognizes assessment as an ongoing and indispensable part of reflective teaching and learning

13.2 Recognizes and understands that assessment must take into account the complex nature of reading, writing, and language and must be based on a range of authentic literacy tasks using a variety of texts

13.3 Is able to conduct assessment that involves a consideration of multiple indicators of learner progress and that takes into account the context of teaching and learning

13.4 Is knowledgeable about the characteristics and appropriate applications of widely used and evolving assessment approaches

13.5 Uses information from norm-referenced tests, criterion-referenced tests, formal and informal inventories, constructed-response measures, portfolio-based assessment, observations, anecdotal records, journals, and multiple other indicators of students; progress to inform instruction and learning

13.6 Recognizes and understands the importance of aligning assessment with curriculum and instruction.

14. Communicating Information About Reading

14.1 Communicates effectively with students, teachers, and support personnel about strengths and areas that need improvement

14.2 Shares pertinent information with other teachers and support personnel

14.3 Understands how to involve parents in cooperative efforts and programs to help students with reading development

14.3 Communicates information about reading programs to administrators, staff members, school board members, parents, and the community

14.4 Effectively communicates information and data about reading to the media, policymakers, and the general public

14.5 Interprets and communicates research findings related to the improvement of instruction to colleagues and the wider community

14.6 Communicates with allied professionals in assessing and planning instruction.

15. Planning and Enhancing Programs—Curriculum and Development

15.1 Initiates and participates in ongoing curriculum development and assessment

15.2 Adapts programs to the needs of different learners to accomplish different purposes

15.3 Supervises, coordinates, and supports all services associated with reading programs (e.g., needs assessment, program development, budgeting and evaluation, grant and proposal writing)

15.4 Understands and uses multiple indicators of curriculum effectiveness.

16. Staff Development

16.1 Initiates, participates in, and evaluates staff development programs

16.2 Takes into account what participants in staff development programs bring to ongoing education

16.3 Provides staff development experiences that help emphasize the dynamic interaction between prior knowledge, experience, and the school context

16.4 Provides staff development experiences that are sensitive to school constraints (e.g., class size, limited resources)

16.5 Understands and uses multiple indicators of professional growth.

17. Research

17.1 Initiates, participates in, or applies researching on reading

17.2 Reads or conducts research within a range of methodologies (e.g., ethnographic, descriptive, experimental, historical)

17.3 Promotes and facilitates teacher- and classroom-based research.

CHAPTER 6

Computer Technology and Reading Instruction

COMPUTER TECHNOLOGY AND READING INSTRUCTION

Executive Summary

Introduction

Although reading is based on the technology of writing and printing, the history of reading instruction reflects a recurrent interest in the application of other technologies, for example, reading pacers, tachistoscopes, and even television. The use of computers in reading instruction dates only to the mid-1960s with the work of Suppes, Atkinson, and their colleagues. For example, Atkinson and Hansen (1966-1967) published the first report of the use of computers in teaching reading. The current review was undertaken to examine research that used computers to deliver reading instruction to determine what the results have been, what the potential is, and what questions remain.

Background

Despite the current intense interest in computer technology, there has been relatively little systematic research into problems of involving computers or other technologies. Several factors seem responsible for the limited research on computers in literacy contexts. First, many reading researchers did not and do not consider technology to be a mainstream topic. That is, they often believe that reading instruction can be delivered *only* by a human. Others believe that technology must be considered in the overall context of reading instruction. Those in the latter category believe that other problems in reading instruction should be attended to before issues of technology are addressed. These general impressions are reinforced by some of the factors described in the following paragraphs.

Until recently, computers did not have all (or even most) of the capabilities that were needed to implement a complete program of reading instruction. A primary lack among these capabilities was the inability to comprehend oral reading and judge its accuracy. Another lack was the inability of computers to accept free-form responses to comprehension questions, leading to reliance solely on recognition tests such as multiple-choice formats. The situation is currently very different, with most new computers capable of speech recognition, as well as a host of multimedia presentation capabilities. Artificial intelligence is beginning to make inroads into software for instruction, and systems for text comprehension are fairly sophisticated, even on home computers.

The development of the Internet and the linking of schools and school computers to it have combined to provide a new interest in computer usage. The kinds of information resources available have provided a stimulus for renewed efforts to deliver instruction of all sorts, including reading, by computer. Coupled with the facts that computers have become much more capable and software has become much more advanced, interest in using the Internet has led to a dramatic new wave of interest in using computers in reading instruction.

Methodology

A database had previously been developed on this topic. That database covered the period from 1986 to 1996 and included all the studies on technology and literacy (e.g., writing as well as reading). Because this database had been developed by a combination of electronic and hand searching all of the journals, it was deemed expedient to use the database and update it with more recent work. Only those studies that met the criteria of the National Reading Panel were included.

Results and Discussion

There is a small body of research on the problems of computer technology and reading. The work that met the National Reading Panel requirements is substantially smaller. Many of the research studies that have been published are explorations of capabilities of computers, comparisons of computers with traditional tasks, word processing, and learning. Very few of these studies directly examined the effects of using computer technology for reading instruction.

A total of 21 studies was found, representing experimental manipulations of problems across the entire spectrum of reading instruction. As a first step to further analysis, the problems addressed by these studies were categorized. The largest group of studies (six) included those that studied the addition of speech to computer-presented text. There were two studies that examined the effects of vocabulary instruction, two more that looked at word recognition instruction, and two that investigated comprehension instruction, broadly defined. One study examined spelling, and two studies examined the effects of broad programs on learning to read. These last studies looked at the delivery of reading instruction by comprehensive software that covered many, if not most, elements of reading instruction.

Conclusions

It is extremely difficult to make specific instructional conclusions based on the small sample of experimental research available. One conclusion is that it is possible to use computer technology for reading instruction. All the studies in the analysis report positive results. The six studies that examined the addition of speech to print presented on computers suggest that this may be a promising alternative, particularly in light of the powerful multimedia computers now available.

There are two other trends that should be examined more systematically. A small, but growing, body of research examines the use of hypertext in learning environments. Although this is technically not reading instruction, it is possible that hypertext might be used in instructional contexts to some advantage.

A second area outside the scope of the current review is that of using computers for word processing. Given that instruction in reading is most efficacious when combined with writing instruction, the use of word processing has the potential to make reading instruction more effective.

Implications for Reading Instruction

Although the Panel is encouraged at the reported successes in using computer technology for reading instruction, there are relatively few specific instructional applications to be gleaned from the research. It is clear that some students can benefit from the use of computer technology in reading instruction. In particular, studies on the addition of speech to print suggest that this may be a promising alternative, especially given the powerful multimedia computers now available and those being developed. In addition, the use of hypertext and word processing appear to hold promise for application to reading instruction.

Directions for Further Research

The reported successes to date in using computer technology for reading instruction indicate that this is an area that needs a great deal of additional exploration. There are many questions that still need to be addressed and many areas in which research does not exist. Particularly striking in its absence is research on Internet applications as they might be incorporated in reading instruction. Another area is the use of computer technology to perform speech recognition. Although great strides have been made in this technology, there have been no recent studies of speech recognition applied to reading instruction, despite its increasing use. Finally, the issue of multimedia presentations has not been addressed in the context of reading instruction. There are many questions that remain about the efficacy of multimedia incorporated in reading instruction.

COMPUTER TECHNOLOGY AND READING INSTRUCTION

Report

Introduction

Reading is based on the technology of writing and printing. The history of reading instruction reflects a recurrent interest in the application of other technologies (Kamil & Lane, 1998; Kamil, Intrator, & Lane, 2000). (The "other technologies" include, for example, reading pacers, tachistoscopes, and television.) The use of computers in reading instruction dates only to the mid-1960s with the work of Suppes, Atkinson, and their colleagues. For example, Atkinson and Hansen (1966-1967) published the first report of the use of computers in teaching reading. This pioneering work in demonstrating the efficacy of using computers to deliver reading instruction set the way for much of the subsequent research. Although there were debates about whether or not they were really teaching reading at the time (Spache, 1968-1969; Atkinson, 1968-1969), such public debates no longer seem to exist.

Despite the current intense interest in computer technology, there has been relatively little systematic research in problems of involving computers or other technologies. Kamil and Intrator (1998) conducted an extensive review of the research in literacy and technology and found that between 1986 and 1996 there were only 350 published research journal articles that reported investigations of reading and writing. The yearly proportion of these technology studies was relatively constant over that time period, ranging from 2% to 5% of the total of all research articles on reading and writing. These totals reflect all research on computers and other technologies, not simply instructional research.

Several factors seem responsible for the limited research on computers in literacy contexts. First, many reading researchers did not and do not consider technology to be a mainstream topic. That is, many believe that reading instruction can be delivered only by a human. Others believe that technology must be considered in the overall context of reading instruction; they believe that other problems in reading instruction should be attended to before issues of technology. These general impressions were reinforced by some of the factors described in the following paragraphs.

Second, for much of the time since the initial reports of computerized reading instruction, computers did not have all (or even most) of the capabilities that were needed to implement a complete program of reading instruction. A primary lack among these capabilities was the inability to comprehend oral reading and judge its accuracy. Another lack was the inability of computers to accept free-form responses to comprehension questions, leading to sole reliance on recognition tests like multiple choice formats.

Lack of those capabilities meant that computer technology often was considered useful only as a supplement to conventional instruction, rather than as a primary delivery system. As a supplemental device, at best, it occupied a less prominent position in the problem space of reading researchers. Indeed, because computer software was relatively incapable of speech recognition or text comprehension, there were only a few activities that the computer seemed to be capable of handling independently. At least in the early history of computers and reading, this was reflected in the translation of things like paper and pencil worksheets to the computer screen. The situation is currently very different, with most new computers capable of speech recognition, as well as a host of multimedia presentation capabilities. Artificial intelligence is beginning to make inroads into software for instruction, and systems for text comprehension are fairly sophisticated, even on home computers.

A third consideration in the history of computers and reading has been the cost factor. With the introduction of microcomputers, the steady decline in prices, accompanied by a steady increase in capabilities, has produced computers that cost only a few hundred dollars. These machines can easily outperform the machines of a decade ago. Most new computers are capable of presenting audio and video, controlling external devices, and being expanded. They have substantial amounts of memory and a great deal of external storage capacity. In addition, there are low-cost printers, scanners, cameras, and a host of other peripherals that can be attached, typically for far less even a few years ago. All of these make unbelievable some of the original predictions that computers would never be cost effective in classrooms.

Finally, there was often resistance to the idea that a computer could deliver reading instruction. One important reason for this simply seems to be the age-old debate about whether teaching reading is an art or a science. Software, to match teacher performance, must be adaptable to a very broad range of responses from learners. It must be able to analyze responses to questions, separating correct from incorrect; respond to idiosyncratic responses in appropriate ways; and bring multiple methods to bear on pedagogical contexts. Computers are still unable to do many of these activities today, despite advances in hardware and software. This problem is not limited to the use of computers to deliver reading instruction. It is endemic to much current software. Despite dramatic developments in learning theory and software design, this seems to be the most serious impediment to progress. The rapid pace of technological innovation in both hardware and software, however, suggests that this issue is being addressed.

At the same time, a different sort of development has caused a renewed interest in computer technology. The development of the Internet and the linking of schools and school computers to it have combined to provide a new interest in computer usage. The kinds of information resources available have provided a stimulus for renewed efforts to deliver instruction of all sorts, including reading, by computer. Coupled with the much greater capability of computers and major advances in software, use of the Internet has led to a dramatic new wave of interest in using computers in reading instruction.

The current review was undertaken to examine the research that used computers to deliver reading instruction in an effort to determine what the results are, what the potential is, and what questions remain.

Methodology

Database

The Technology Subcommittee began its task by using a database that was assembled by Kamil and Intrator (1997). This database was deemed an appropriate starting point because it was assembled by a combination of electronic searches and exhaustive hand searches of all the journals that appeared in the electronic searches. The following paragraphs describe the methods used in the creation of the database in that study.

A review of the research on computer technology and reading was undertaken to document the trends in this area. To accomplish this task, the first step was to interrogate both the ERIC and PsycINFO databases. Any journal research article that matched the descriptors of technology, computers, reading, writing, or literacy was listed.

Queries were generated in the form of SUBJECT READ# and SINCE 1986 not YEAR = 1996 and DOCTYPE = research and DOCTYPE = journal article and S = technology. Other queries were composed to cover similar topics in reading, writing, speaking, listening, and literacy. Both "technology" and "computer" were used as qualifiers. The Panel decided that single descriptors would yield a more liberal sampling of articles, even though such a procedure yields more "false positives." (For example, using literacy as a descriptor yielded many studies of science literacy and computer literacy that did not deal with reading or writing.) These queries yielded a total of 965 articles in 159 different journals.

In a preliminary hand search of the journals, evidence was found that there were articles that did not appear in the ERIC or PsycINFO databases. Consequently, each of the 159 journals was hand-searched for relevant articles that were missed or missing from the database interrogation. In addition, many of the articles in the original set did not meet the criteria of true research reports about literacy and technology. For example, some of the articles were mere speculation; others were about computer literacy rather than reading. Still others were commentaries arguing for or against the efficacy of technology interventions in literacy learning. The Panel applied a simple criterion to include or exclude articles. To be included, an article had to deal with the areas laid out above and had to be based on an empirical data collection. However, reviews of research studies based on empirical studies were included. Because the original search was conducted prior to the end of 1996, the Panel included any available 1996 issues of journals in the hand search.

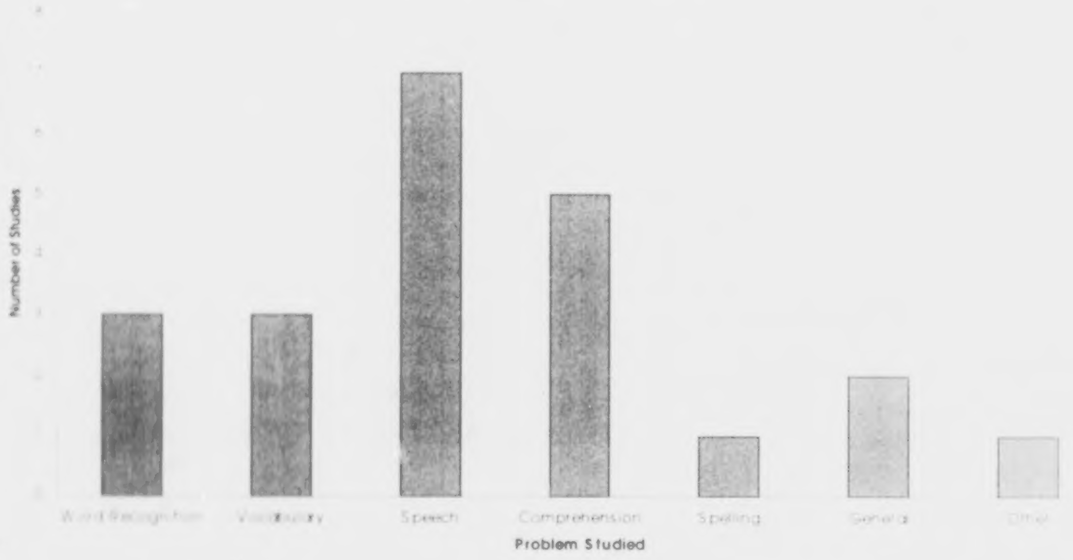
Subsequent additions to the database were made by queries of the INSPEC database and hand-searching similar to that described above. This yielded an ultimate pool of 350 studies. Information on all relevant articles was entered into a Filemaker Pro database. Each article was assigned a value for number of pages, literacy type, technology type, subject population, special population characteristics, problem, platform, methodology, findings, recommendations, and quality.

For the present analysis, the studies in the database were filtered to identify a subset of experimental or quasi-experimental instructional studies. Of the 350 studies, a total of 92 investigated reading using experimental or quasi-experimental methods. Of the 92, only 47 studies were instructional. Studies that merely compared computerized versions of a task with conventional versions were not counted as instructional. Studies that merely examined effects of the computer, unless attended by some instructional intervention, were also eliminated from the pool. What this last criterion did was to remove a few studies that simply translated existing materials for use in a computer presentation. Studies that did not deal with computer technologies were also eliminated. (In the original database, for example, there were studies that examined instructional uses of television.) Studies that were about word processing were not considered further, because many or most of these did not involve any connection with reading. Finally, studies of special populations, non-native speakers of English, or adult readers were deemed inappropriate for further analysis. (A number of studies dealt with learning to read in a second language, for example, and fell outside the scope of the charge to the National Reading Panel.) This produced a final pool of only 21 studies.

Analysis

The 21 studies represent experimental manipulations of problems across the entire spectrum of reading. As a first step to further analysis, the problems addressed by these 21 studies were categorized. This procedure ultimately yielded seven categories. The largest group of studies (six) studied the addition of speech to computer-presented text. There were two studies that examined the effects of vocabulary instruction, two more that looked at word recognition instruction, and two that investigated comprehension instruction, broadly defined. One study examined spelling, and two studies examined the effects of broad programs on learning to read. The last studies looked at the delivery of reading instruction by comprehensive software that covered

Figure 1. Number of Computer Technology Studies as a Function of Reading Problems (N = 22 Problems in 21 Studies)



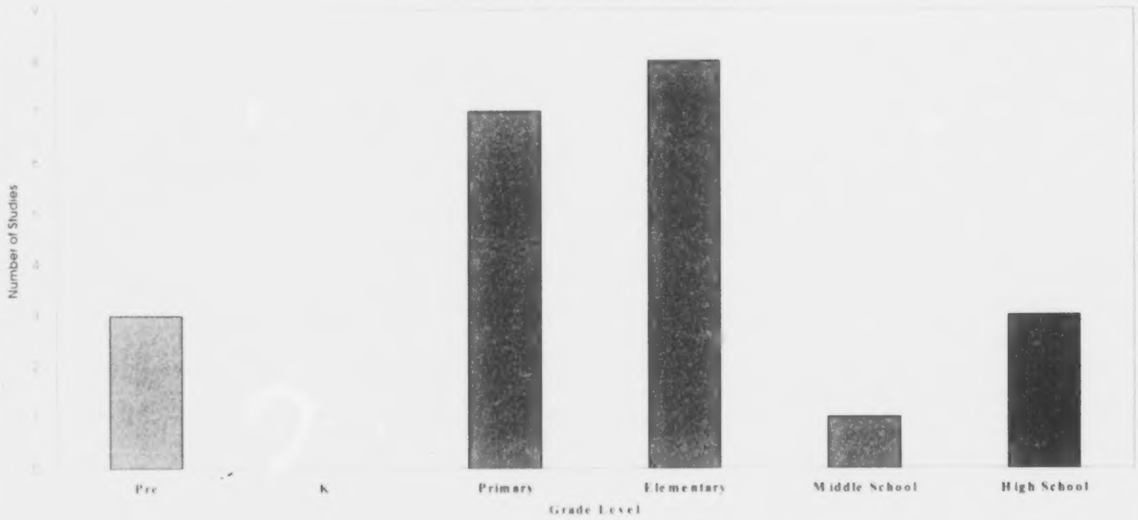
BEST COPY AVAILABLE

many, if not most, elements of reading instruction. One study examined the learning of picture-word relationships and was not classified. Figure 1 presents these data graphically.

Consistency With NRP Methods

Meta-analysis was judged inappropriate because the 21 studies were spread across the entire spectrum of variables and across populations ranging from preschool to high school. The distribution of the final pool of studies as a function of grade levels is included as Figure 2. What is interesting about this distribution is that it is equally focused on primary and elementary students. The implication is that technology has been applied with equal emphasis to problems of early readers and of more experienced ones. Perhaps the anomaly is that there are so few studies at the high school level.

Figure 2. Distribution of Studies as a Function of Grade Level of the Learners
(N=16 Studies with 23 Grade Samples)



A striking feature of the entire pool of studies is the small proportion of studies that used experimental or quasi-experimental methods compared to the total number of instructional studies. Only 92 of the 350 studies, or 26%, used experimental methods. Moreover, fewer than 5% of the studies in the original data set met the criteria for inclusion established by the National Reading Panel. Even though these numbers are low, they are in agreement with the conclusions of Kamil and Intrator (1997), Kamil and Lane (1998), and Kamil, Intrator, and Kim (2000) that there has been a dearth of research on problems in technology and literacy. According to Kamil and Intrator (1997), there was no significant increase in published research on technology and literacy over the time span from 1986 to 1996.

BEST COPY AVAILABLE

Results

It is difficult to conclude much on the basis of the 21 studies. They all report successful uses of computer technology in one context or another. Kamil and Intrator (1997) classified the studies in their database according to whether the processes studied were old or new modes of instruction. For example, an "old" process might be completing a workbook page at the computer rather than with paper and pencil. A new one might be reading from hypertext.

They further classified the old modes as to whether they merely replaced an old form of instruction or augmented it. If, for example, the workbook page was merely completed and the student was given no feedback, this was a simple replacement. If, however, the student was given appropriate instruction, based on the answers, it was considered to be an augmented process.

In the final data set of instructional studies, there were no new processes studied. They were equally divided between the augmented or replacement categories. This seems to suggest that, for the experimental research, there are few examples of truly new uses for computer technology to date.

This is an important finding in that it suggests that there are few truly innovative uses of computer technology in literacy instruction, despite the great promise. There will almost certainly be more developments of new uses for technology in literacy instruction in the future. For now, the computer seems to be used as technology to either present or augment traditional instructional practices.

Discussion

There are threads in the research database that are worth noting even if the studies on which they are based do not meet the formal criteria established by the NRP. These findings are based on a limited amount of data, and not all of these studies are purely instructional. They are given here to indicate that there are factors not quite central to reading instruction that might be adapted for such use. Before strong recommendations could be made that these should be incorporated in reading, additional research would be needed. These trends include the potential benefits of computers in reading for word processing, use of computers as motivational devices, use of computers as assistive technologies, and the potential of hypertext as an alternative medium for reading and studying. These trends are consistent with the trends noted by Kamil, Intrator, and Kim (2000).

In particular, the database contains 131 studies (38%) that were about writing. Although not all of these were instructional studies, a number were. They were, however, excluded, as noted above, by the formal criteria established by the NRP. The exclusion of these studies is not meant to imply that the teaching of writing is unimportant. The Panel believes it can be integrated with beginning reading instruction in beneficial ways. What was missing from the published research was an explicit link to reading instruction. Consequently, these studies were not included.

There are reviews of the specific literature on word processing that already exist, and it was considered unproductive to duplicate them. The Panel suggests that the use of word processing in writing instruction could be an important and effective addition to the reading curriculum that can benefit immediately from the use of computer technology.

A second cluster of studies involved the use of computer technologies as motivational agents. The Panel judged that these studies were, again, not directly in the instructional charge but worth considering. It is probably the case that as computers become more familiar to students, their motivational value will diminish. For the present, though, this still seems to be a potent variable, although its precise application is far from clear. Again, reading instruction can probably make good use of the motivational aspects of computers and software.

The third trend is reflected in a set of studies that examines what Kamil, Intrator, and Kim (2000) have called *assistive* technologies. There were 114 studies in the original database (33% of the total) that dealt with special populations. While not all of these are experimental or quasi-experimental studies, they do point to an important cluster of research activities. There seems to have been less resistance to the adoption of computer technologies for these populations than for mainstream populations. Although less evidence is presented of the effectiveness of computers for use in mainstream instructional applications, the uses with special populations may point the way for the future.

Finally, the NRP looks to the promise of hypertext as an application for the future. A small, but steadily increasing, cluster of studies points the way toward potentially important applications revolving around hypertext and hypermedia. There were 12 studies that involved hypertext in the assembled database, but they do not adequately reflect the growing interest in the topic. Many of the studies do not meet the experimental or instructional criteria, but they will provide important data as researchers and practitioners conceptualize new ways to apply hypermedia in reading and learning to read. It will be those applications that must be researched and validated for use in reading instruction. Hypertext and hypermedia may also involve developing new modes of instruction for students to use them effectively. What is most exciting about this trend is that it represents truly new ways of applying computer technology to reading and reading instruction.

Implications for Reading Instruction

There are few implications for practice that can be drawn from the small set of instructional studies in the database. What is important is that there are uses for the computer that do impinge on reading instruction. The following is an attempt to draw some of these implications, with the caveat that the implications are clearly tentative and need to be verified by continued research.

Computers can be used for some reading instructional tasks.

Although there are only a few experimental studies that are relevant to this point, they do report successes. What is clear is that as computer software becomes more capable, the opportunities for computers to be used in reading instruction will expand.

At the very least, computers can provide opportunities for students to interact instructionally with text for greater amounts of time than they can if only conventional instruction is provided. Although there was no research that provided a general rule for determining what works, careful selections from available software can provide additional instructional assistance in classrooms. Although there is a publication bias to report only positive differences, there were no instructional studies in which the computer did not provide a significant addition to the instructional context.

Word processing is a useful addition to reading instruction.

A very large portion of the database involved studies of word processing. Because writing is often part of reading instruction, the findings concerning word processing are relevant, even though the studies fell outside the criteria for analysis. Word processing has many benefits for writing, particularly in its close match with process writing approaches. Although the implication has not been experimentally tested (in terms of its effect on reading instruction), this seems likely to occur in the future. One implication seems to be that word processing alone is unlikely to make a difference; it must be embedded in other instruction.

- **Multimedia computer software can be used for reading instruction.**

There are many unanswered questions about the efficacy of multimedia learning. All of the conditions under which multimedia learning is more effective than conventional learning are not known. However, there appear to be many students who benefit from the addition of multimedia instruction to a conventional curriculum. One example that was tested in several studies was the addition of speech (computerized or not) to the instructional context. When multimedia software is available and appropriate, it should be exploited.

- **Computers do have a motivational use in reading instruction.**

Although there were no experimental instructional studies that supported this implication explicitly, the motivational aspects of computers should not be overlooked. This effect may diminish as computers become ever more common. For the time being, they still retain some motivational advantage over conventional instruction.

Hypertext has a great deal of potential in reading instruction.

There is a growing interest in hypertext because of its potential to allow the reader to control some of the presentation of text, determining what to read at various junctures in the text. Another potential is the use of hypertext to assist the reader who is having difficulty with a passage. Despite the fact that there were no experimental instructional studies on this topic that met the NRP criteria, the application of hypertext concepts to reading and reading instruction seems to have a great deal of potential. The use of hypertext and hypermedia on the Internet almost mandates the need to address this issue in reading instruction. In the meantime, hypertext, particularly coupled with Internet access, seems to have been adopted in many classrooms, regardless of the lack of research.

Directions for Further Research

It is inappropriate to separate the applications of computer technologies to reading instruction into a set of issues about technology and a set about reading. The Panel believes that technology is not a problem to be studied in and of itself. The problem is, rather, how the technology is applied to specific problems in reading instruction. To that end, the following questions are offered as among the most important ones to be answered by future research. They are neither simple nor easily answered. Answering them will involve issues as complex as professional development for teachers and as simple as the utility of drill and practice exercises.

Research on these topics needs to be relatively independent of specific computer platforms and software because the rapidity of innovation makes specific choices obsolete in short time periods. One argument for not conducting more research has been that the technology outpaces the research. However, not all of the important questions are dependent on the state of technological innovation.

The Panel believes that the following list of questions represents relatively short-term needs for today and shortly beyond the horizon of current development. Some effort should be directed at conceptualizing new uses for computer technology—uses that will augment conventional reading instruction in beneficial ways. The list does not include questions that may become important in the future—such as the role of literacy in a much more graphically oriented world. These may not be researchable, but the implications of these developments need to be systematically explored by research.

One of the most striking findings of this analysis is that there is a surprising lack of published research. For whatever reason, the volume of published research has not kept pace with the interest in computer technology. Research is urgently needed to answer these and other questions that will affect the penetration of computer technology into conventional reading instruction.

1. What is the proper role for integration of computers in reading instruction? In what contexts can they be used to either replace or supplement conventional instruction?
2. What are the conditions under which multimedia presentation is useful or desirable in reading text?
3. What are the requisite characteristics of software to teach reading?
4. What is the appropriate mix of reading and writing instruction delivered by computer?
5. How can professional development programs be structured to help teachers effectively integrate computer solutions with instruction?
6. How are the effects of computer usage in pedagogy most effectively measured? Do conventional assessments measure all of the learning that takes place in computer environments?
7. What is the utility of hypertext in instructional contexts?
8. How can Internet resources be incorporated in reading instruction?

Overall Conclusions

The current analysis has found general agreement in the experimental literature that computer technology can be used to deliver a variety of types of reading instruction successfully. There has been relatively little research in this important area and, consequently, many unanswered questions remain.

The rapid development of capabilities of computer technology, particularly in speech recognition and multimedia presentations, promises even more successful applications in literacy for the future. To be certain that these new developments are incorporated in instruction as efficiently as possible, it is important that research be initiated to answer the questions that have not been addressed to date.

References

Atkinson, R. (1968-1969). A reply to reaction to computer-assisted instruction in initial reading: The Stanford Project. Reading Research Quarterly, 3, 418-420.

Atkinson, R., & Hansen, D. (1966-1967). Computer-assisted instruction in initial reading: The Stanford Project. Reading Research Quarterly, 2, 5-26.

Barker, A. B., & Torgeson, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. Journal of Educational Computing Research, 13, 89-103.

Bass, G., Ries, R., & Sharpe, W. (1986). Teaching basic skills through microcomputer-assisted instruction. Journal of Educational Computing Research, 2, 207-219.

Chambless, J., & Chambless, M. (1994). The impact of instructional technology on reading/writing skills of 2nd grade students. Reading Improvement, 31, 151-155.

Chang, L., & Osguthorpe, R. (1990). The effects of computerized picture-word processing on kindergartners' language development. Journal of Research in Childhood Education, 5, 73-84.

Davidson, J. (1994). The evaluation of computer-delivered natural speech in the teaching of reading. Computers and Education, 22, 181-185.

Davidson, J., Coles, D., Noyes, P., & Terrell, C. (1991). Using computer-delivered natural speech to assist in the teaching of reading. British Journal of Educational Technology, 22, 110-118.

Davidson, J., Elcock, J., & Noyes, P. (1996). A preliminary study of the effect of computer-assisted practice on reading attainment. Journal of Research in Reading, 19, 102-110.

Gillingham, M., Garner, R., Guthrie, J., & Sawyer, R. (1989). Children's control of computer-based reading assistance in answering synthesis questions. Computers in Human Behavior, 5, 61-75.

Gore, D., Morrison, G., Maas, M., & Anderson, E. (1989). A study of teaching reading skills to the young child using microcomputer assisted instruction. Journal of Educational Computing Research, 5, 179-85.

Icabone, D., & Hannaford, A. (1986). Comparison of two methods of teaching unknown reading words to fourth graders: Microcomputer and tutor. Educational Technology, 26, 36-39.

Kamil, M. L., & Intrator, S. (1997, December). Quantitative trends in publication of research on technology and reading, writing, and literacy. Scottsdale, AZ: The National Reading Conference.

Kamil, M. L., & Intrator, S. (1998). Quantitative trends in publication of research on technology and reading, writing, and literacy. In T. Shanahan & F. Rodriguez-Brown (Eds.). Forty-seventh Yearbook of the National Reading Conference (pp. 385-396). Chicago, IL: The National Reading Conference.

Kamil, M. L., Intrator, S., & Kim, H. S. (2000). Effects of other technologies on literacy and literacy learning. In M. Kamil, P. Mosenthal, P. D. Pearson, et al. (Eds.) Handbook of reading research (Vol. 3) (pp. 773-788). Mahwah, NJ: Lawrence Erlbaum Associates.

Kamil, M. L., & Lane, D. (1998). Researching the relationship between technology and literacy: An agenda for the 21st century. In D. Reinking, M. McKenna, L. Labbo, et al. (Eds.). Handbook of literacy and technology: Transformations in a post-typographic world. (pp. 323-341). Mahwah, NJ: Lawrence Erlbaum.

Kolich, E. (1991). Effects of computer-assisted vocabulary training on word knowledge. Journal of Educational Research, 84, 177-182.

Norton, P., & Resta, V. (1986). Investigating the impact of computer instruction on elementary students' reading achievement. Educational Technology, 26, 35-41.

Olson, R. K., Foltz, G., & Wise, B. W. (1986). Reading instruction and remediation with the aid of computer speech. Behavior Research Methods, Instruments, and Computers, 18, 93-99.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. Reading Research Quarterly, 23, 484-498.

Reinking, D., & Rickman, S. S. (1990). The effects of computer-mediated texts on the vocabulary learning and comprehension of intermediate grade readers. Journal of Reading Behavior, 22, 395-411.

Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening, and independent reading with computer-based speech feedback. Reading Research Quarterly, 23, 219-235.

Roth, S., & Beck, I. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. Reading Research Quarterly, 22, 197-218.

Spache, G. (1968-1969). A reaction to computer-assisted instruction in initial reading: The Stanford Project. Reading Research Quarterly, 3, 101-109.

Tobias, S. (1988). Teaching strategic text review by computer and interaction with student characteristics. Computers in Human Behavior, 4, 299-310.

Weber, W., & Henderson, E. (1989). A computer-based program of word study: Effects on reading and spelling. Reading Psychology, 10, 157-171.

Wise, B. (1992). Whole words and decoding for short-term learning: Comparisons on a talking-computer system. Journal of Experimental Child Psychology, 54, 147-167.

Wise, B., Olson, R., & Treiman, R. (1990). Subsyllabic units in computerized reading instruction: Onset-rime vs. postvowel segmentation. Journal of Experimental Child Psychology, 49, 1-19.

Appendix A

Studies Analyzed

Barker, A. B., & Torgeson, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. Journal of Educational Computing Research, 13, 89-103.

Bass, G., Ries, R., & Sharpe, W. (1986). Teaching basic skills through microcomputer assisted instruction. Journal of Educational Computing Research, 2, 207-219.

Chambless, J., & Chambless, M. (1994). The impact of instructional technology on reading/writing skills of 2nd grade students. Reading Improvement, 31, 151-155.

Chang, L., & Osguthorpe, R. (1990). The effects of computerized picture-word processing on kindergartners' language development. Journal of Research in Childhood Education, 5, 73-84.

Davidson, J. (1994). The evaluation of computer-delivered natural speech in the teaching of reading. Computers and Education, 22, 181-185.

Davidson, J., Coles, D., Noyes, P., & Terrell, C. (1991). Using computer-delivered natural speech to assist in the teaching of reading. British Journal of Educational Technology, 22, 110-118.

Davidson, J., Elcock, J., & Noyes, P. (1996). A preliminary study of the effect of computer-assisted practice on reading attainment. Journal of Research in Reading, 19, 102-110.

Gillingham, M., Garner, R., Guthrie, J., & Sawyer, R. (1989). Children's control of computer-based reading assistance in answering synthesis questions. Computers in Human Behavior, 5, 61-75.

Gore, D., Morrison, G., Maas, M., & Anderson, E. (1989). A study of teaching reading skills to the young child using microcomputer assisted instruction. Journal of Educational Computing Research, 5, 179-185.

Icabone, D., & Hannaford, A. (1986). Comparison of two methods of teaching unknown reading words to fourth graders: Microcomputer and tutor. Educational Technology, 26, 36-39.

Kolich, E. (1991). Effects of computer-assisted vocabulary training on word knowledge. Journal of Educational Research, 84, 177-182.

Norton, P., & Resta, V. (1986). Investigating the impact of computer instruction on elementary students' reading achievement. Educational Technology, 26, 35-41.

Olson, R. K., Foltz, G., & Wise, B. W. (1986). Reading instruction and remediation with the aid of computer speech. Behavior Research Methods, Instruments, and Computers, 18, 93-99.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. Reading Research Quarterly, 23, 484-498.

Reinking, D., & Rickman, S. S. (1990). The effects of computer-mediated texts on the vocabulary learning and comprehension of intermediate grade readers. Journal of Reading Behavior, 22, 395-411.

Reitsma, P. (1988). Reading practice for beginners: Effects of guided reading, reading-while-listening, and independent reading with computer-based speech feedback. Reading Research Quarterly, 23, 219-235.

Roth, S., & Beck, I. (1987). Theoretical and instructional implications of the assessment of two microcomputer word recognition programs. Reading Research Quarterly, 22, 197-218.

Tobias, S. (1988). Teaching strategic text review by computer and interaction with student characteristics. Computers in Human Behavior, 4, 299-310.

Weber, W., & Henderson, E. (1989). A computer-based program of word study: Effects on reading and spelling. Reading Psychology, 10, 157-171.

Wise, B. (1992). Whole words and decoding for short-term learning: comparisons on a talking-computer system. Journal of Experimental Child Psychology, 54, 147-167.

Wise, B., Olson, R., & Treiman, R. (1990). Subsyllabic units in computerized reading instruction: Onset-rime vs. postvowel segmentation. Journal of Experimental Child Psychology, 49, 1-19.

MINORITY VIEW

MINORITY VIEW

Joanne Yatvin, Ph.D.

Oregon Trail School District, Sandy, Oregon

The charge from Congress to the National Reading Panel (NRP) was to "assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read." In explicating that charge, the National Institute of Child Health and Development (NICHD), which convened the Panel, listed seven questions for the Panel to address. They were:

1. What is known about the basic processes by which children learn to read?
2. What are the most common instructional approaches in the United States to teach children to learn to read? What are the scientific underpinnings for each of these methodologic approaches, and what assessments have been done to validate their underlying scientific rationale? What conclusions about the scientific basis for these approaches does the Panel draw from these assessments?
3. What assessments have been made of the effectiveness of each of these methodologies in actual use in helping children develop critical reading skills, and what conclusions does the Panel draw from these assessments?
4. Based on the answers to the preceding questions, what does the Panel conclude about the readiness for implementation in the classroom of these research results?
5. How are teachers trained to teach children to read, and what do studies show about the effectiveness of this training? How can this knowledge be applied to improve this training?
6. What practical findings from the Panel can be used immediately by parents, teachers, and other educational audiences to help children learn to read, and how can the conclusions of the Panel be disseminated most effectively?
7. What important gaps remain in our knowledge of how children learn to read, the effectiveness of different instructional methods for teaching reading, and improving the preparation of teachers in reading instruction that could be addressed by additional research?

From this charge, it seems reasonable to infer that Congress's goal was to settle the "Reading Wars," putting an end to the inflated rhetoric, partisan lobbying, and uninformed decisionmaking that have been so widespread and so detrimental to the progress of reading instruction in America's schools. Clearly, the main thrust of the charge is toward determining which of the many teaching methods used in schools, and promoted by advocates, really work best.

Whether a review of the existing reading research literature could have provided answers to all of Congress's questions, the Panel's obligation was to dig in and find out. I am filing this minority report because I believe that the Panel has not fulfilled that obligation. From the beginning, the Panel chose to conceptualize and review the field narrowly, in accordance with the philosophical orientation and the research interests of the majority of its members. At its first meeting in the spring of 1998, the Panel quickly decided to examine research in three areas: alphabetics, comprehension, and fluency, thereby excluding any inquiry into the fields of language and literature. After some debate, members agreed to expand their investigations to two other areas: computer-linked instruction and teacher preparation. Five subcommittees were formed, and within the chosen areas, each selected a number of topics of interest. As work on the initial choices of topics proceeded, however, it became apparent that the Panel had insufficient time and support personnel to cover all it had identified. Ultimately, the Panel subgroups produced reviews of the research on the following topics: phonemic awareness, phonics, fluency, comprehension strategies, vocabulary development, computer technology and reading instruction, teacher preparation in general, and teacher preparation to teach comprehension strategies. In addition, the Panel developed a set of criteria and procedures for evaluating reading studies, which all subgroups used and which the Panel hopes will serve as future guidelines for other researchers.

These reviews show comprehensive and painstaking work by the subcommittees. They will prove valuable, I think, to other experimental researchers as they seek to expand the body of knowledge on those topics and fill in the gaps. On the other hand, the reviews are of limited usefulness to teachers, administrators, and policymakers because they fail to address the key issues that have made elementary schools both a battleground for advocates of opposing philosophies and a prey for purveyors of "quick fixes." And, unfortunately, the reviews are of even less use to parents because they do not touch on early learning and home support for literacy, matters which many experts believe are the critical determinants of school success or failure.

To have properly answered its charge, the Panel had to look at the field of reading both horizontally and vertically, examining the basic theoretical models of reading, the methods that grow out of them, and the processes of learning that begin in infancy and continue through young adulthood. (See Appendix A for definitions and descriptions of the three models underlying methods of instruction in American schools today.) The scientific basis for each of these models needed to be examined, then the effectiveness of the methods they have generated. The research on language development, pre-reading literary knowledge, understanding of the conventions of print, and all the other experiences that prepare young children to learn to read also demanded the Panel's attention. And finally, the changing needs and strategies of adolescent readers called for a review of the existing research.

If the Panel could not cover the whole field—as, in fact, it could not because of time and resource limitations—it should have concentrated on topics of highest interest and controversy in the public arena. Or, as professionally distasteful as the task might have been, it should have assessed the validity of the claims of various commercial programs being sold as cure-alls to schools and parents. (In order to be specific about topics the panel did not cover, I have included two lists in Appendix B.) The panel chose not to pursue any of these approaches.

Furthermore, to have fully answered its charge, the Panel needed to assess the implications for practice growing out of research findings. As a body made up mostly of university professors, however, its members were not qualified to be the sole judges of the "readiness for implementation in the classroom" of their findings or whether the findings could be "used immediately by parents, teachers, and other educational audiences." Their concern, as scientists, was whether or not a particular line of instruction was clearly enough defined and whether the evidence of its experimental success was strong. What they did not consider in most cases were the school and classroom realities that make some types of instruction difficult—even impossible—to implement. Outside teacher reviewers should have been brought in to critique the Panel's conclusions, just as outside scientists were to critique its processes. Despite repeated suggestions that this be done, it was not.

In fairness to the Panel, it must be recognized that the charge from Congress was too demanding to be accomplished by a small body of unpaid volunteers, working part time, without staff support, over a period of a year and a half. (The time Congress originally allotted was only 6 months.)

Congress did not realize—and the Panel itself did not fully comprehend at the beginning of its labors—how large, uneven, and intractable the field of reading research really is. The Panel's preliminary electronic searches of databases uncovered thousands of articles on some topics, hundreds on others, only a handful on some. Their completed reviews on several topics disclosed that the critical question of generalizability (i.e., Does a skill or strategy taught and learned carry over to new experiences?) often was not answered by researchers. The reviews show, in addition, that questions relevant to the success of an instructional technique, such as "how much" to teach and "when," were not even examined in most studies.

501

Also in fairness to the Panel, I must acknowledge that a few of the topics I have identified as neglected are included in some of the reports. Still, they receive only peripheral attention when public interest demands much more. In the review on phonemic awareness, for example, the critical question of whether all children need special training in phonemic awareness was not addressed, even though several studies suggest that many children grasp the concept and are able to apply it through ordinary reading instruction. Other topics of interest, such as students' need for "direct instruction," appear in reviews only as assumptions about successful practices, but are never tested against their philosophical opposites.

In the end, the work of the NRP is not of poor quality; it is just unbalanced and, to some extent, irrelevant. But because of these deficiencies, bad things will happen. Summaries of, and sound bites about, the Panel's findings will be used to make policy decisions at the national, state, and local levels. Topics that were never investigated will be misconstrued as failed practices. Unanswered questions will be assumed to have been answered negatively. Unfortunately, most policymakers and ordinary citizens will not read the full reviews. They will not see the Panel's explanations about why so few topics were investigated or its judgments that the results of research on some of the topics are inconclusive. They will not hear the Panel's calls for more and more fine-tuned research. Ironically, the report that Congress intended to be a boon to the teaching of reading will turn out to be a further detriment.

As an educator with more than 40 years of experience and as the only member of the NRP who has lived a career in elementary schools, I call upon Congress to recognize that the Panel's majority report does not respond to its charge nor meet the needs of America's schools. In spite of the Panel's diligent efforts and its valuable findings on a select number of instructional practices, we still cannot answer the first and most central question of the charge: "What is known about the basic processes by which children learn to read." We still do not know what types of instruction are suitable for different ages and populations of children. We still do not know the relative effectiveness of the three models of reading as bases for instruction. We do not even know whether the existing body of research can answer those questions. Therefore, I ask Congress not to take actions that will promote one philosophical view of reading or constrain future research in the field on the basis of the Panel's limited and narrow set of findings.

Appendix A

Definitions

Word Identification Model of Reading

The word identification model hypothesizes that readers read by matching letters to sounds, then blending sounds into pronounceable words. In asserting that children who have mastered the skills of decoding "can read anything," it separates word pronunciation from word understanding and defines the former as reading. Instructional materials evade the issue by using mostly decodable words in stories that reflect familiar life experiences of children and have only literal meanings.

Although proponents of this model recognize that readers need vocabulary knowledge and skills of analysis and interpretation to understand advanced and specialized materials, they believe that the job of developing those skills properly belongs in subject matter classes. Getting students to understand the main idea of a short story, for example, is the business of the literature teacher, not the reading teacher, and is better left to middle and high school grades.

This model does not consider the factor of reader motivation. At all levels the reader is viewed as a passive recipient of content. Children should learn to read because adults want them to. They should remember the facts in a text and accept the teacher's interpretation of meaning. Because of these beliefs, there are few attempts to make reading an interesting or rewarding experience for children.

Word Identification Plus Skills Model of Reading

In this model, learning to read is a two-tier process. The first tier is very much like that of the previous model, except that it defines reading as understanding words as well as pronouncing them. Children are able to read sentences, paragraphs, and whole texts by stringing together the pronunciations and meanings of individual words.

The second tier of the process is "reading to learn." As readers gain speed and automaticity in recognizing words and verbalizing sentences naturally, they free up their mental abilities to deal with larger vocabulary loads and implied meanings. However, because this model, like the first, views readers as recipients of content, they need direct instruction in comprehension strategies. Through instruction, readers learn how to deal with different kinds of texts and their increasing length, complexity, and subtlety.

Reader motivation is a part of this model, but it is seen mostly as an external factor: What must the teacher do to move children to read this story and do the accompanying activities?

Integration of Language and Thinking Model of Reading

According to this model, children begin acquiring the knowledge and skills needed for reading long before they face the challenge of decoding print. Even at the earliest stages of reading, they are able to use what they know about language, literature, and the world to perform multiple operations in dealing with a text. Reading means not only recognizing words and knowing their meanings, but also understanding how they fit into a context of grammatical structure, speech phrasing and intonation, literary forms and devices, and print conventions.

Because readers bring their own skills and knowledge to any text, and because written language is redundant, they are able to orchestrate their own reading experiences. When one skill or knowledge source is weak in relation to a particular text, such as life experience would be in reading about the history of a foreign country, stronger skills, such as vocabulary, may carry the reader through. In this model, learning to read and reading to learn are inseparable.

Although this model also recognizes the need for reader strategies in dealing with more difficult texts, it views strategies as the products of individual needs and purposes, sometimes devised by the reader and sometimes prompted or provided by others at the point of need. Motivation, then, needs to be intrinsic. The teacher's job is to create or allow situations where children want to read and are willing to work hard at it.

Learning to read in this model involves "others" in many ways. Readers expand their vocabularies and background knowledge through listening to the teacher read stories aloud and conversing with their peers. They adopt and adapt strategies modeled by others. They modify their understanding of texts by listening to what others have to say. At the same time, roles continually change: the questioner is questioned, and the explainer is corrected. Thus, social interaction is a necessary component of this model.

Appendix B

Below are two lists of topics not investigated by the National Reading Panel. The first is drawn from a survey of leaders in reading from across the United States done by the International Reading Association (Reading Today, December 1999). These leaders were asked to identify what topics they perceived to be "hot" in the field today. The second list is my own view of topics that teachers and parents are concerned about, either because they are now in wide use or are being advocated for inclusion in the reading curriculum.

International Reading Association List of "Hot" Topics

- Balanced reading instruction
- Decodable text
- Direct instruction
- Early intervention
- Performance assessment
- Standards
- State/national assessment
- Volunteer tutoring

My List of Topics of Public Concern

- Direct instruction
- Use of decodable texts
- Embedded skills instruction
- Reading aloud to children
- Invented spelling
- Use of predictable texts
- Early language development (vocabulary, grammar, and literary language)
- Integrated reading and writing
- Home-teaching programs
- Access to quality literature
- Whole-class instruction
- Scripted instruction
- Teacher modeling
- Children's understanding of print conventions

Appendix C

Dear Panel Members:

I spent most of Friday and yesterday at the annual conference of the Oregon Reading Association. Although I was not scheduled to speak, I was introduced at the first general session as a member of the National Reading Panel (NRP). Because of that introduction, I was later approached by a number of teachers who thanked me for representing them and who expressed the hope that the Panel's report would relieve the pressure from the state legislature and local school boards to adopt one-sided commercial programs that would take away their authority to decide what is best for their students and that would consume most of the time allocated for reading over several years of schooling. I did not have the heart to tell them that the NRP Report would probably open the door to increased pressure rather than lessen it.

I was also engaged in conversation by two reading researchers who testified at the Panel's regional meeting in Portland in 1998. They called then for the inclusion of ethnographic research in the Panel's investigations and have since learned that it was not included. They could not see any logic or fairness in that decision. I did not tell them that their appeals at the Portland meeting and those of like-minded colleagues at other regional meetings were not even mentioned in the Panel's Executive Summary.

In addition, I attended a presentation by Patricia Edwards, a member of the International Reading Association (IRA) Board, who has done research on the effects of home culture on children's literacy development. She did not have to persuade me; this area of early language development and literary and world experience is the one I believe is most critical to children's school learning, and the one I could not persuade the Panel to investigate. Without such an investigation, the NRP Report's coverage of beginning reading is narrow and biased.

Over the past 2 months, I have wavered about whether it was useful or right for me to submit a minority report. I waver no longer. I hereby reiterate my request that the minority report I submitted in January and include in this e-mail (with minor revisions), be sent to Congress along with the majority report. Only in that way can I honorably serve the teachers and children I represent.

Joanne Yatvin

February 27, 2000

U.S. Department of Health and Human Services
Public Health Service
National Institutes of Health
National Institute of Child Health and Human Development
NIH Pub. No. 00-4754
April 2000

506